MCB
Journals.ASM.org

Dynamic Switching of Active Promoter and Enhancer Domains Regulates *Tet1* and *Tet2* Expression during Cell State Transitions between Pluripotency and Differentiation

Abhishek Sohni,[a] Michela Bartoccetti,[a] Rita Khoueiry,[a] Lien Spans,[b] Joris Vande Velde,[a] Linde De Troyer,[a] Kirthi Pulakanti,[c] Frank Claessens,[b] Sridhar Rao,[c] Kian Peng Koh[a]

KU Leuven Department of Development and Regeneration, Stem Cell Institute Leuven, Leuven, Belgium[a]; KU Leuven Department of Cellular and Molecular Medicine, Laboratory of Molecular Endocrinology, Leuven, Belgium[b]; Blood Center of Wisconsin, Blood Research Institute, Milwaukee, Wisconsin, USA[c]

The Tet 5-methylcytosine dioxygenases catalyze DNA demethylation by producing 5-hydroxymethylcytosine and further oxidized products. *Tet1* and *Tet2* are highly expressed in mouse pluripotent cells and downregulated to different extents in somatic cells, but the transcriptional mechanisms are unclear. Here we defined the promoter and enhancer domains in *Tet1* and *Tet2*. Within a 15-kb "superenhancer" of *Tet1*, there are two transcription start sites (TSSs) with different activation patterns during development. A 6-kb promoter region upstream of the distal TSS is highly active in naive pluripotent cells, autonomously reports *Tet1* expression in a transgenic system, and rapidly undergoes DNA methylation and silencing upon differentiation in cultured cells and native epiblast. A second TSS downstream, associated with a constitutively weak CpG-rich promoter, is activated by a neighboring enhancer in naive embryonic stem cells (ESCs) and primed epiblast-like cells (EpiLCs). *Tet2* has a CpG island promoter with pluripotency-independent activity and an ESC-specific distal intragenic enhancer; the latter is rapidly downregulated in EpiLCs. Our study reveals distinct modes of transcriptional regulation at *Tet1* and *Tet2* during cell state transitions of early development. New transgenic reporters using *Tet1* and *Tet2 cis*-regulatory domains may serve to distinguish nuanced changes in pluripotent states and the underlying epigenetic variations.

The study of transcription regulation is fundamental to understand how gene expression and phenotype are regulated during development. ENCODE (encyclopedia of DNA elements) studies recently revealed that the mammalian genome is more complex than previously annotated, in which different transcription start sites (TSSs) often mark core promoters, both intra- and intergenic, to drive the expression of alternative mRNA isoforms (1, 2). Distal elements known as enhancers are often recruitment platforms for cell-type-specific transcription factors and interact with promoter-bound factors to stabilize the association of RNA polymerase II (Pol II) at TSSs (3).

Embryonic stem cells (ESCs) derived from the inner cell mass (ICM) of mouse blastocysts represent a unique model of transcription regulation. The widespread presence of "open" chromatin achieved through feed-forward and autoregulatory loops involving transcription factors appears crucial for the transcriptome to adapt rapidly to inductive differentiation signals, allowing the cells to generate all early embryonic lineages (4, 5). This pluripotent state is dependent on the master transcription factors Oct4, Sox2, and Nanog (OSN), which often bind cooperatively at target sites on enhancers. In early mammalian development, the triad of OSN, by binding to different enhancers, orchestrates two states of pluripotency, the "naive" and "primed" conditions resembling the pre- and postimplantation epiblasts, respectively (6, 7). Furthermore, OSN together with Klf4, Esrrb, and Mediator coactivator, all highly expressed in naive ESCs, densely occupy extended domains of enhancer clusters, named "superenhancers," which are associated with genes that play major roles in ESC identity (8). During differentiation, superenhancer-associated genes are highly sensitive to reduced Mediator occupancy and are rapidly silenced.

Among the list of superenhancer-associated genes in ESCs are *Tet1* and *Tet2*, encoding DNA-modifying enzymes (8). Together with *Tet3*, the TET dioxygenases act on 5-methylcytosine (5mC), the preeminent mark of DNA methylation in the mammalian genome, to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) (9–11). The reiterative oxidation steps constitute both passive and active pathways in DNA demethylation; at the genome scale, this process is a fundamental part of epigenetic reprogramming (12). *Tet1* and *Tet2* are highly expressed in mouse ESCs and primordial germ cells relative to somatic cells: both have been implicated in early embryonic differentiation (13, 14), primordial germ cell specification (15, 16), imprinting (17, 18), and induced pluripotency (19, 20). In the adult, *TET2* shows broader tissue-specific expression than does *TET1* (21). *Tet3*, although not part of the ESC gene signature, is expressed highly in many tissues, including oocytes, where it contributes to DNA hydroxymethylation in the paternal genome early postfertilization (22). The loss of 5hmC and *TET* expression

Received 18 September 2014 **Returned for modification** 28 October 2014
Accepted 30 December 2014

Accepted manuscript posted online 12 January 2015

Citation Sohni A, Bartoccetti M, Khoueiry R, Spans L, Vande Velde J, De Troyer L, Pulakanti K, Claessens F, Rao S, Koh KP. 2015. Dynamic switching of active promoter and enhancer domains regulates *Tet1* and *Tet2* expression during cell state transitions between pluripotency and differentiation. Mol Cell Biol 35:1026–1042. doi:10.1128/MCB.01172-14.

Address correspondence to Kian Peng Koh, kian.koh@med.kuleuven.be.

A.S. and M.B. contributed equally.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/MCB.01172-14.

doi:10.1128/MCB.01172-14

is also a hallmark of many cancers (23, 24). Despite many studies on *TET* using loss- and gain-of-function approaches, the transcriptional mechanism underlying their tissue-specific expression remains to be clarified.

We previously showed that *Tet1* and *Tet2* expressions are dependent on *Oct4* and *Sox2* (13). Using mouse-human sequence conservation to predict regulatory elements, we identified conserved Oct4 sites in *Tet1* and *Tet2*. A subsequent study suggests that Oct4, but not Sox2, is required for *Tet2* transcription by acting at the conserved Oct4-Sox2 motif (25). In this study, we identified the TSS of *Tet1* by 5′ rapid amplification of cDNA ends (5′ RACE). Using previously reported high-coverage chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-seq) data sets, we defined promoter and enhancer regions in *Tet1* and *Tet2*, including additional elements in the superenhancer domains, and validated their cell state-specific activities by cell-based reporter assays. In addition, we examined the contribution of Oct4 and Sox2 motifs within the enhancer domains to transcription factor occupancy and gene expression. Finally, we examined DNA methylation changes at specific regulatory regions during early pluripotency states and differentiation.

MATERIALS AND METHODS

Cell culture. The mouse v6.5 ESC line (C57BL/6 × 129/sv F1 hybrid) was cultured in standard ESC medium on feeder layer as previously described (13). Cells were harvested by trypsin-EDTA treatment and replated twice on tissue culture wells over 30 to 45 min to remove the adherent feeders from the ESC in suspension. Culture in defined medium containing dual MEK and glycogen synthase kinase (GSK) inhibitors (2i medium and differentiation toward epiblast-like cells (EpiLCs) were performed as previously described (6). The epiblast-derived stem cell (EpiSC) lines FT.1, FT.2, and FT.3 (26) were a generous gift from Alice Jouneau. NIH 3T3 (a kind gift from Francois Fuks) and HEK293T cells were cultured under standard conditions in high-glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with GlutaMAX and 10% fetal bovine serum (FBS) (Sigma-Aldrich).

Mouse embryos and adult tissues. Embryos were isolated by mechanical dissection of the decidua isolated from the uterine lining of naturally mated C57BL/6 mice. Epiblasts from mice at embryonic day 6.0 postcoitum (E6.0) to E7.0 were extracted by first dissecting the embryos free of Reichert's membrane and peripheral trophectodermal tissues with the aid of fine tungsten needles. Next, the embryonic visceral endoderm was removed by incubating the embryos for 10 min at 4°C in 2.5% pancreatin–0.5% trypsin solution (in Ca^{2+}/Mg^{2+}-free Tyrode Ringer's saline). Finally, the epiblasts were separated from all other extraembryonic tissues. Each epiblast sample was pooled from 3 to 5 embryos for RNA or genomic DNA extraction. Embryos from E8.5 to E10.5 were dissected free of yolk sac and amnion. Mouse embryonic fibroblasts (MEFs) were derived from E13.5 mouse embryos and cultured for 2 to 3 passages before harvest. Adult tissues were dissected from 6- to 8-week-old female mice. All experimental procedures on mice have been reviewed and approved by the KU Leuven Ethical Committee for Animal Experimentation (P177/2011) in compliance with the European Directive 2010/63/EU.

5′ rapid amplification of cDNA ends. Total RNA was isolated by using TRIzol for 5′ RACE as according to the manufacturer's instructions (Invitrogen). Briefly, cDNA was synthesized with SuperScript II reverse transcriptase by using a gene-specific primer (see Table S1 in the supplemental material). A 3′ homopolymeric cytosine or adenine tail was added by using terminal deoxynucleotidyltransferase (TdT) for PCR amplification. Purified PCR products were cloned into TOPO 2.1 (Invitrogen) for sequencing.

Quantitative reverse transcription-PCR (RT-PCR). Total RNA was extracted from tissues following homogenization in TRIzol (Invitrogen) according to the manufacturer's instructions. RNA was purified from cells or from pooled epiblasts/individual embryos by using the RNeasy minikit or microkit, respectively (Qiagen), with on-column DNase I digestion. cDNA was synthesized by using SuperScript III (Invitrogen). Quantitative PCR (qPCR) was performed by using SYBR green PCR master mix (Invitrogen) on a StepOnePlus or ViiA7 real-time PCR system (Life Technologies). Absolute standard curves were generated by 10-fold serial dilutions of plasmid clones containing target sequences, starting from 10^5 or 10^6 copies estimated based on mass concentration. Transcript copy numbers were determined by calibration to standard curves and normalized by *Gapdh* gene copy numbers.

ChIP-seq analysis. ChIP-seq Sequence Read Archive (SRA) files were obtained from the GEO (see Table S2 in the supplemental material). Reads were aligned to mm9 by using the command line (-e 70 -k 1 -m 1 -n 2 -best -concise) on the software BowTie. Peaks were then identified by using MACS (model-based analysis of ChIP-seq) (66), using a two-sided comparison with input DNA when available, with default parameters (27).

Reporter plasmid construction. Genomic fragments of *Tet1* and *Tet2* were amplified from the bacterial artificial chromosome (BAC) genomic clones RP23-132J10 and RP24-333H9, respectively, by using KOD DNA polymerase (Novagen). Primers for subcloning are listed in Table S1 in the supplemental material. Putative promoter fragments were subcloned into the pGL3-Basic vector (Promega) at either the MluI/XhoI or KpnI/XhoI sites. Putative enhancer fragments were subcloned into the pGL-3-Promoter vector (Promega), containing a minimal simian virus 40 (SV40) promoter (or endogenous *Tet1* and *Tet2* promoter fragments), at the downstream SalI site in either the sense or antisense orientation to the luciferase gene (*Luc*). Methylated plasmids were generated *in vitro* by treating 10 to 18 μg of plasmids with 80 U CpG methyltransferase M.SssI (New England BioLabs) in NEBuffer2 containing 640 μM *S*-adenosylmethionine for 4 h at 37°C. Complete methylation was verified by protection against cleavage by HpaII and by bisulfite sequencing of promoter regions.

Site-directed mutagenesis. High-performance liquid chromatography (HPLC)-purified primers containing sequence modifications were used for PCR with KOD DNA polymerase (Novagen). The product was subjected to DpnI (Promega) treatment to digest the parental vector. The linear vector was gel purified, phosphorylated by using T4 polynucleotide kinase (Thermo Scientific), ligated by using T4 DNA polymerase (New England BioLabs), and transformed into competent bacteria. Modified fragments were verified by sequencing and further subcloned into the pGL3-Promoter vector(s).

Dual-Luciferase reporter assay. On the day of transfection, v6.5 ESCs were seeded at 1×10^5 to 2×10^5 cells per well of a 12-well plate containing feeder cells for transfection using Lipofectamine 2000 (Invitrogen). To transfect cells during the ESC-to-EpiLC conversion, 2i ESCs were seeded at 0.7×10^5 to 1×10^5 cells per well of a 12-well plate on day 0 in EpiLC differentiation medium and transfected as adherent cells on day 1; as controls, ESCs were seeded at similar densities in standard serum medium with leukemia inhibitory factor (LIF) on gelatin-coated plates and transfected in parallel. NIH 3T3 cells were seeded at 1×10^5 to 2×10^5 cells per well of a 6-well plate and transfected the next day as adherent cells by using TransIT-3T3 (Mirus Bio). All pGL3 constructs were cotransfected with pRL-TK at a ratio of 10:1 to normalize the transfection efficiency. Cells were harvested 2 days after transfection (or 36 h after the ESC-to-EpiLC conversion) for measurement of luciferase activities by using a Dual-Luciferase reporter assay system (Promega). Data are means ± standard errors of the means (SEM) from at least 3 independent transfections, each performed in duplicate.

Electrophoretic mobility shift assay. Nuclear extracts (NEs) from ESCs or HEK293T cells were prepared from 2×10^7 to 5×10^7 cells, as previously described (28). Oligonucleotides with 5′ overhangs containing guanine(s) were hybridized, filled in with [α-^{32}P]dCTP using Klenow fragments, and purified on a Sephadex column as labeled probes. A total

of 5 to 10 μg of the NE was incubated with the probe (20,000 cpm) on ice for 20 min in 1× binding buffer (20 mM HEPES [pH 7.9], 0.1 mM EDTA, 5 mM $MgCl_2$, 17% [vol/vol] glycerol, 100 mM NaCl) containing 50 ng/μl poly(dI·dC) (Sigma-Aldrich), 20 mM dithiothreitol (DTT), and 1% Triton X. Supershift was performed by adding the antibodies for Sox2 (sc-17320) (28) and Oct4 (sc-5279) (Santa Cruz Biotech) to the NEs in binding buffer for 20 min, followed by the addition of labeled probe and incubation on ice for an additional 20 min. Similarly, unlabeled (cold) competitor duplex oligonucleotides were added at a 100× to 300× molar excess.

Bisulfite sequencing. Genomic DNAs were purified and bisulfite treated by using an EpiTect bisulfite conversion kit (Qiagen). Pooled epiblast specimens were processed by using the EpiTect Plus LyseAll bisulfite kit (Qiagen). Tet-assisted bisulfite (TAB) sequencing was performed by using the 5hmC TAB-Seq kit according to the manufacturer's protocol (WiseGene). PCR products were generated by using Hifi *Taq* polymerase (Invitrogen), gel purified, and subcloned into the pGEM-T Easy vector (Promega) for sequencing. The CpG methylation status of the sequences was analyzed using QUMA (67). Samples with a conversion rate of <95% and a sequence identity of <90%, as well as identical bisulfite sequences, were excluded. Primers for bisulfite PCR are listed in Table S1 in the supplemental material.

Generation of a fluorescence reporter transgenic line. The pGmch2p construct was derived from the pGL3 vector by removal of the *Luc* gene between the HindIII and XbaI sites and replacement with a bicistronic cassette of mCherry and a puromycin resistance open reading frame (ORF) linked by a viral 2A peptide (mCherry-2A-PuR). Thirty micrograms of purified, NotI-linearized plasmid was electroporated into 1×10^7 v6.5 ESCs at 320 V and 250 μF by using a Gene Pulser XCell system (Bio-Rad). Electroporated cells were plated onto puromycin-resistant feeders and treated 24 h later with 1.5 μg/ml puromycin for 7 to 10 days to select for resistant colonies. mCherry-positive clones were picked and propagated under standard ESC culture conditions. Cellular fluorescence was monitored and imaged with a Zeiss Axiovert 40 CFL microscope. Cell suspensions were collected for flow cytometry on a MACSQuant VYB instrument and analyzed by using FlowJo software.

Statistical analysis. The significance of mean differences was calculated by analysis of variance (ANOVA) and a multiple-comparison *post hoc* test using GraphPad Prism. *P* values of <0.05 were considered statistically significant.

Accession numbers. Clonal sequences obtained by 5′ RACE analyses of Tet1 as shown in Fig S1 in the supplemental material have been deposited at the European Nucleotide Archive under accession numbers LN810022 to LN810047 and are accessible on at www.ebi.ac.uk/ena/data/view/LN810022-LN810047. The NCBI Reference Sequence (RefSeq) number for the *Tet1* mRNA transcript as shown in the UCSC genome browser is NM_001253857.1. The genome build of NCBI *Mus musculus* Annotation Release 104 contains additional mRNA reference sequences predicted by automated computational analysis: the 5′ untranslated region (UTR) starting with exon 1b as described in this study matches sequences under accession numbers XM_006513873.1, XM_006513866.1, XM_006513868.1, and XR_380416.1; the 5′ UTR starting with the exon 1a sequence as annotated in this study matches a sequence under accession number XM_006513867.1; a variant exon 1a sequence (found in 5′ RACE clones 7 and 10 as shown in Fig. S1E) matches the sequence under accession number XM_006513869.1.

RESULTS

Identification of two cell state-dependent TSSs in *Tet1*. The NCBI reference murine *Tet1* mRNA transcript lacks a 5′ untranslated region (UTR). To determine the TSS of *Tet1*, we performed 5′ RACE with mouse v6.5 ESCs (see Fig. S1 in the supplemental material). All sequenced clonal products had in common a 119-bp 5′ UTR contiguous with the coding sequence (CDS), constituting exon 2, but spliced to different noncoding exons (exon 1) upstream. The majority (21 of 28 clones) of 5′ ends aligned with a site 12.7 kb upstream of exon 2, here denoted exon 1b (Fig. 1A; see also Fig. S1 in the supplemental material). A few clones (4 of 28) mapped to a site 7.2 kb upstream, denoted exon 1a, which shares sequence similarity with noncoding exon 1 annotated in human *TET1* (Fig. 1A). The remaining clones (3 of 28) mapped to different sites further upstream of exon 1b but, as they were minor products, were not investigated further (see Fig. S1E in the supplemental material). Both exons 1a and 1b aligned with strong binding peaks of the initiating form of Pol II (Pol2_5P) in the mouse ESC genome (29), consistent with both sites having TSSs. Binding peaks for the elongation form (Pol2_2P) suggest active transcription across the gene body downstream of exon 1a. Weak Pol II peaks were also detected in the vicinity of the Ensembl noncoding exon (ENSMUSE00000930068) located 18.8 kb upstream of the CDS, which we denote exon 1c (Fig. 1A), although this transcript was never detected by 5′ RACE.

To measure the relative abundance of these 5′-UTR transcript variants, we performed quantitative RT-PCR (qPCR) using a forward primer specific for each noncoding exon (exons 1a, 1b, and 1c) paired with a common reverse primer in exon 2 (Fig. 1A; see also Fig. S2A in the supplemental material). All three spliced 5′-UTR variants were detected in mouse ESCs by RT-PCR (see Fig. S2B in the supplemental material). Levels of expression of 5′-UTR variants were compared to that from a coding region (detected by using a primer pair specific for exons 12 and 13) in terms of transcript copy numbers. Based on these measurements, the TSS at exon 1b appeared to be the most active; transcript copy numbers detected at exon 1b exceeded those detected at the 3′ coding region and those at exon 1a by 2-fold and 6-fold, respectively (2.1- ± 0.3-fold and 6.0- ± 0.8-fold) ($n = 6$) (Fig. 1B). Transcripts starting from exon 1c were hardly detectable (see Fig. S2C in the supplemental material). The greater abundance of 5′-end transcripts than of full-length copies suggests a steady-state level of transcriptional initiation exceeding elongation at the *Tet1* TSS, which is indicative of regulation at the postinitiation step. A genome-wide analysis using human cells described transcription initiation at most genes, only a subset of which experiences elongation, as a general phenomenon (30).

We observed simiilar transcript profiles for *Tet1* in feeder-cultured mouse ESCs and induced pluripotent stem cells (iPSCs) (Fig. 1B). To determine whether this profile reflects heterogeneity within ESC cultures, we compared feeder-free cultures of ESCs maintained with exogenous leukemia inhibitory factor (LIF) in the presence of serum (serum ESCs) and in defined "2i" medium (2i ESCs); the latter condition is thought to promote a more homogeneous population of naive cells in a "ground" state of pluripotency (31). The depletion of feeders from v6.5 cells resulted in a 2- to 3-fold diminution in the overall level of *Tet1* transcripts, associated with increased spontaneous differentiation, but the levels of exon 1b consistently exceeded those of exon 1a by 5- to 6-fold (5.5- ± 0.5-fold) ($n = 3$) and were comparable to the levels in the feeder-independent ESC lines CJ7 and E14Tg2a grown in serum (Fig. 1C; see also Fig. S2D in the supplemental material). Culture in 2i, which sustains cells as undifferentiated rounded colonies, caused a modest reduction in overall *Tet1* expression levels compared to those in serum ESCs but retained the relative exon 1b/1a expression ratio of about 5 (4.8 ± 0.7):1 ($n = 6$) (Fig. 1C). The latter finding is in agreement with previous studies describing modest or variable differences in *Tet1* expression between

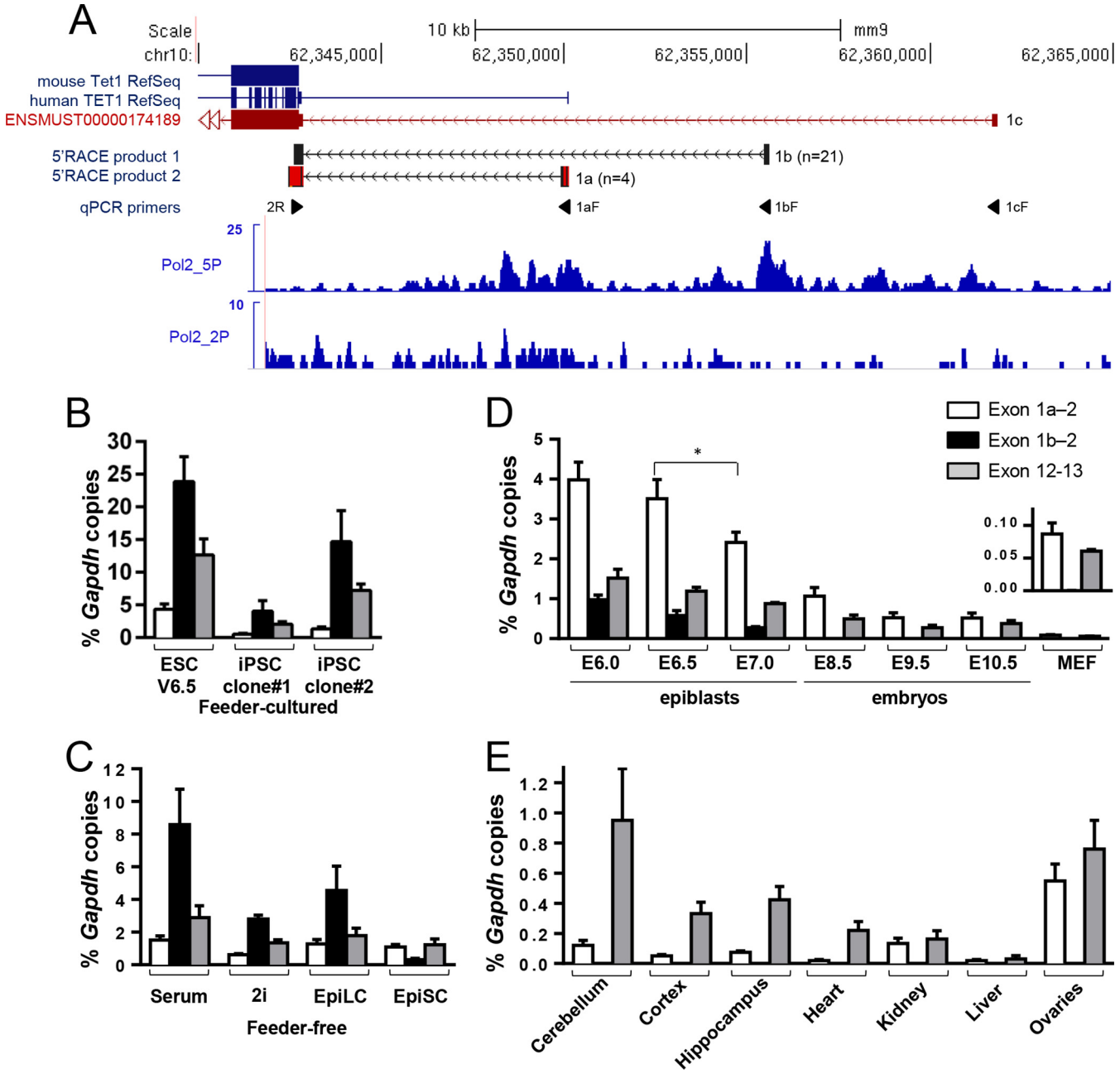


FIG 1 Identification of an ESC-specific TSS in *Tet1*. (A) Alignment of 5′-RACE clonal products from v6.5 ESCs to *Tet1* (mm9) by using the BLAST-like alignment tool (BLAT) shown in the UCSC genome browser view. The number of clones sequenced is indicated in parentheses (see Fig. S1 in the supplemental material). An Ensembl transcript is also shown, with exon 1 labeled 1c. The positions of qPCR detection primers are indicated (see Fig. S2A in the supplemental material). At the bottom, Integrative Genome Viewer (IGV) tracks of ChIP-seq occupancy of the initiating (Pol2_5P) and elongating (Pol2_2P) forms of Pol II in ESCs (29) are shown. (B to E) Normalized transcript copy numbers detected within the *Tet1* 5′ UTR (exons 1a and 1b) and CDS (exons 12 and 13) in feeder-cultured v6.5 ESCs and iPSCs (B), in 2i- and serum-cultured v6.5 ESCs in the presence of LIF versus EpiLCs and EpiSCs (C), in mouse embryonic tissues (D), and in tissues of adult mice (E). The inset in panel D shows the results for MEFs, with an enlarged *y* axis. Data are means ± SEM from >3 biological replicates. *, $P < 0.05$.

2i and serum conditions (32–34). Although we cannot formally rule out that individual cells may selectively activate one TSS over the other without single-cell analysis, our results suggest that the two TSSs are more likely to be concurrently active in each cell.

To investigate further whether *Tet1* expression changes during early postimplantation development, we used a recently reported protocol to differentiate 2i ESCs into epiblast-like cells (EpiLCs) in 48 h (35). The converted EpiLCs displayed a gene expression profile similar to that of early, pre-primitive-streak stage postimplantation embryos, as previously described (6), representing a state of primed pluripotency (see Fig. S2E in the supplemental material). In EpiLCs, the *Tet1* transcript profile remained similar to that in 2i ESCs, although the ratio of exon 1b/1a decreased to 3.3 ± 0.5 ($n = 4$). In contrast, mouse epiblast-derived stem cells

(EpiSCs), which were recently shown to resemble the late-gastrula-stage epiblast/ectoderm upon *in vitro* adaptation (36), showed an altered profile, in which exon 1a transcripts were sustained but exon 1b transcripts were further diminished to low levels (with a ratio of exon 1b to 1a transcripts of 0.3 ± 0.1 to 1 $n = 3$) (Fig. 1C).

To correlate these observations with changes *in vivo*, we examined native epiblasts of E6.0 to E7.0 early gastrulating mouse embryos and whole embryos at E8.5 to E10.5 postgastrulation stages. In prestreak stage (E6.0) and early-streak stage (E6.5) mouse epiblasts, we observed the expression of exon 1a transcripts at ESC levels (3.7% ± 0.3% of *Gapdh* levels; $n = 6$) and already diminished levels of exon 1b transcripts (<1% of *Gapdh* levels) (Fig. 1D). The difference between EpiLCs *in vitro* and E6.0 epiblasts *in vivo* suggests that EpiLCs may correlate more closely to the E5.0 to E5.5 epiblasts not examined in this study. Levels of exon 1a and 1b transcripts declined in mid-streak-stage (E7.0) embryos, becoming more similar to the transcript profile of EpiSCs. By E8.5, exon 1b was no longer detected in whole embryos. TSS at exon 1a sustained low levels of *Tet1* expression in E8.5 to E10.5 embryos (0.3 to 1.6% of *Gapdh* levels) but was further downregulated in primary cultures of E13.5 mouse embryonic fibroblasts (Fig. 1D). In the adult mouse, low levels of *Tet1* (<1% of *Gapdh* levels) persisted in all tissues examined; of the three 5′ transcript variants examined (exons 1a, 1b, and 1c), only exon 1a was detectable (Fig. 1E). Slightly elevated levels of *Tet1* coding transcripts were observed in tissues of the brain (cerebellum, cortex, and hippocampus), as previously described (37, 38), and in the ovaries. Quantitations were similar using primers specific for either exons 4 and 5 or exons 12 and 13 (data not shown). However, in brain tissues and also in the heart, the exon 1a 5′ transcripts consistently did not account for the higher copy numbers of transcripts detected in the coding region, suggesting that there are alternative 5′ transcripts not detected by our 5′-RACE and qPCR assays (Fig. 1E). 5′ RACE performed by using mouse cerebellum cDNA confirmed only one TSS at exon 1a (data not shown). One possible explanation for this finding is that transcription initiation occurs in the dispersed mode over the promoter region upstream of exon 1a. Sequences of 5′-RACE clonal products in ESCs that aligned close to exon 1a (see Fig. S1E in the supplemental material) indeed suggest TSSs at variant positions within the 200-bp vicinity of the exon 1a sequence annotated in this study (see Fig. S1F in the supplemental material), some of which cannot be detected by our qPCR primers.

Collectively, these results suggest a strong preference for TSS at exon 1b of *Tet1* in the preimplantation ICM and early epiblast, which rapidly switches to a preference for a TSS at exon 1a in the postimplantation epiblast. In late-postgastrula-stage embryos and in the adult soma, only the TSS at exon 1a is weakly active.

Cell state-specific coupling of promoter and enhancer domains in *Tet1*. To identify *cis*-regulatory elements in *Tet1*, we surveyed its epigenome in mouse ESCs using ChIP-seq data sets (Fig. 2A; see also Table S2 in the supplemental material). The combinatorial binding profiles of pluripotency factors, Mediator, and cohesin units have previously delineated a superenhancer domain in *Tet1*, a 15.2-kb region that spans from exon 1c at the distal end to a conserved noncoding sequence (CNS) upstream of the CDS (8) (Fig. 2A; see also Fig. S3 in the supplemental material). ChIP-seq showed strong binding profiles of Oct4 within the CNS, as previously shown (13), but also several more strong peaks in the distal region between exons 1b and 1c (Fig. 2A). We also examined specific chromatin signatures of histone H3 lysine 4 trimethylation (H3K4me3), which marks active promoters, and histone H3 lysine 4 monomethylation (H3K4me1) in combination with the presence of histone H3 lysine 27 acetylation (H3K27ac) and the absence of H3K4me3, which frequently marks active enhancers (39, 40). Consistent with exons 1a and 1b having active TSSs, both sites were associated with strong and broad H3K4me3 marks (Fig. 2A). Signals for H3K4me1 were detected across the entire gene locus, but the signals for H3K27ac were highly enriched over a 22.9-kb region, which encompasses and extends upstream of the superenhancer. In ESCs, the gene locus is entirely devoid of signals for the silencing mark H3K27me3 (data not shown).

Based on these chromatin signatures and binding patterns of *trans*-acting factors, we defined two genomic segments as putative promoter regions: the 5-kb segment A and 6-kb segment B upstream of and including exons 1a and 1b, respectively. These segments were subcloned as 5′-serially deleted fragments upstream of a reporter gene (Fig. 2A). Segment B contains 6 prominent ChIP-seq peaks that are common binding sites for OSN factors, p300, and Mediator (see Fig. S3 in the supplemental material); therefore, 6 subfragments were tested, which represent truncations to progressively remove each ChIP-seq peak region (Fig. 2A). The constructs were transfected into either mouse v6.5 ESCs or NIH 3T3 cells (a murine fibroblast line, as a nonpluripotent control) for *in vitro* reporter assays.

Within genomic region A, all 5 fragments of various lengths showed no promoter activity in ESCs and low basal activity in NIH 3T3 cells (see Fig. S4A in the supplemental material). In contrast, all fragments within segment B, except for the shortest 0.5-kb fragment, fragment B1, showed strong promoter activity in ESCs at least 10-fold ($n = 4$ to 6) above basal levels observed for NIH 3T3 cells, suggesting that the promoter activity within region B is highly ESC specific (Fig. 2B). The 3.4-kb fragment B4 (base positions −3242 to +168 of TSS at exon 1b) consistently showed the strongest ESC-specific activity, significantly exceeding the activities of all other fragments in ESCs. Adjustment of the transfection ratios of plasmids to account for their different sizes had no apparent effect on their relative activity within the limits of the transfection conditions (see Fig. S4B in the supplemental material). To determine whether the differences were due to the presence of enhancer or repressor elements, we tested the individual 1-kb fragments D1 to D4, each compassing an OSN peak in region B for regulatory activity. In reporter constructs driven by either a minimal SV40 promoter or the endogenous *Tet1* promoter (fragment B3), all D subfragments caused no effect or only a modest 2- to 3-fold increase in promoter activity (see Fig. S4C and D in the supplemental material). These experiments showed no evidence for any strong distal-acting enhancer or repressor constituents within the D fragments. Alternatively, these regions may function as *cis* regulators only in the context of the entire 6-kb promoter. Thus, we regard the entire 6-kb fragment B6 as a functionally autonomous full promoter. Fragment B1, with only one OSN peak region, may constitute a minimal proximal promoter.

We next tested the 1.4-kb fragment E, which compasses the CNS and an Oct-Sox (O/S) motif, for enhancer function (Fig. 2A). In either orientation, fragment E increased SV40 promoter activity by 5- to 15-fold ($n = 5$) in ESCs (significantly higher in the sense [>10-fold] than in the antisense [5- to 10-fold] direction) and not at all in NIH 3T3 cells, which is indicative of a highly ESC-specific enhancer (Fig. 2C). We further examined the effect

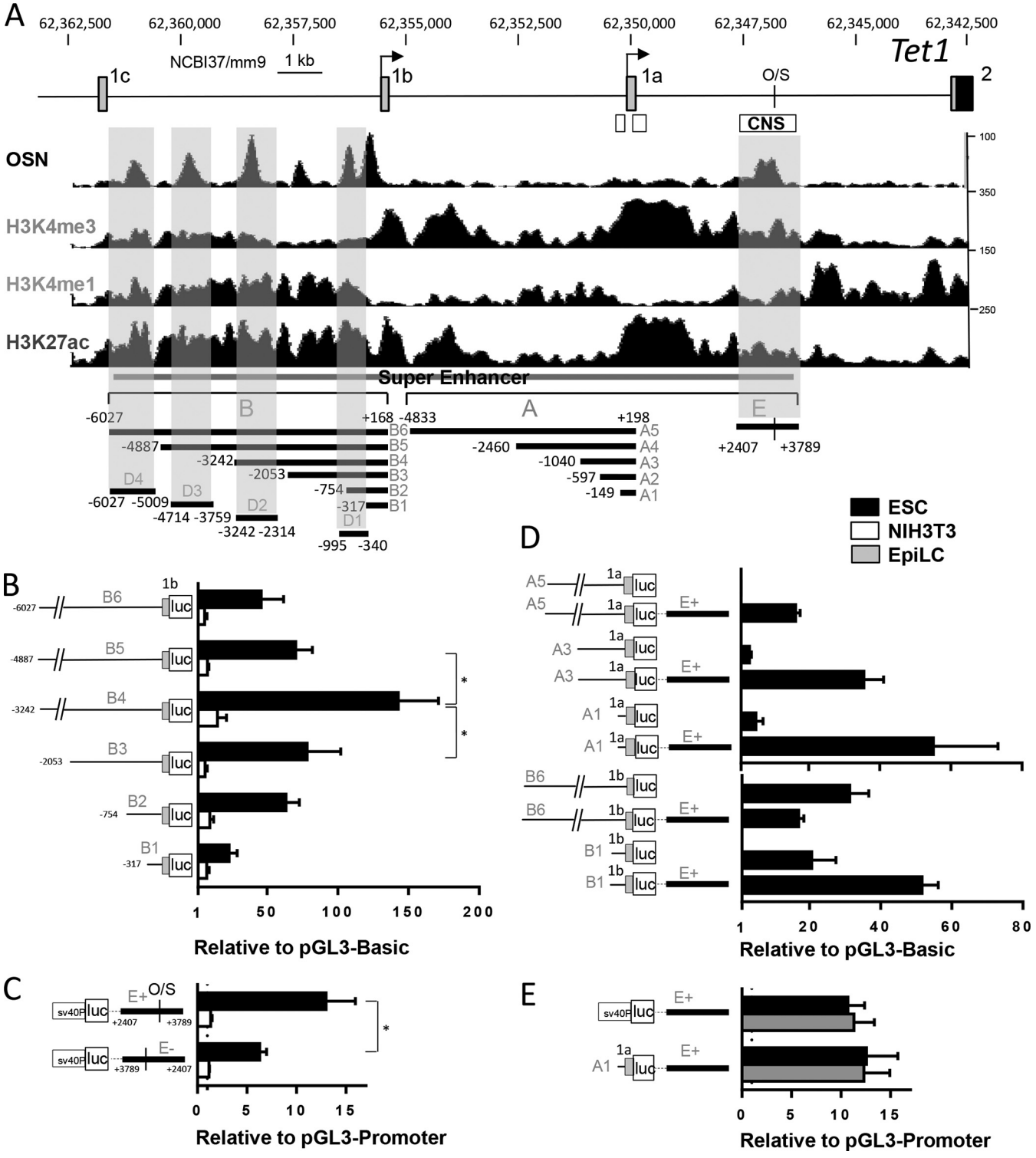


FIG 2 Characterization of *Tet1* regulatory domains by luciferase assays. (A) Binding profiles of OSN (represented by Oct4 ChIP-seq) (8) and histone H3 modifications (65) over the locus upstream of *Tet1* in ESCs. The 5′-UTRs are denoted as gray boxes, and the CDS in exon 2 is denoted as a black box. ChIP-seq signals are shown as tag densities with the *y* axis floor set to 1. The locus is shown in reverse compared to Fig. 1A. The CNS (white boxes) is based on mouse-human conservation and an Oct-Sox (O/S) motif, indicated with a vertical line. The superenhancer domain is marked by a gray line. At the bottom, bars mark fragments tested, with base positions at the ends denoted with respect to TSS at exon 1a for fragments A and E and with respect to TSS at exon 1b for fragments B and D. (B to E) Normalized luciferase expression of constructs (shown by figures along the *y* axes) to test for promoter (B and D) or enhancer (C and E) activities relative to those of control vectors (set to 1 and shown as dotted lines in panels C and E) in ESCs, NIH 3T3 cells, and EpiLCs. Data are means ± SEM from >3 biological replicates. *, $P > 0.05$.

of fragment E cloned in the sense direction downstream of endogenous *Tet1* promoter sequence A or B. Interestingly, fragment E activated sequence A fragments of various sizes, which all showed negligible promoter activity by themselves *in vitro* (see Fig. S4A in the supplemental material), in ESCs but not in NIH 3T3 cells (Fig. 2D and data not shown). Fragment E enhanced the activity of fragment B1 by a variable 3.8- ± 1.4-fold ($n = 4$) but had no stimulatory effect on the full promoter B6 (Fig. 2D). Thus, fragment E is *cis* enhancer that regulates *Tet1* promoter region A in ESCs but can be dispensable for promoter region B. To determine whether fragment E shows similar enhancer activities in EpiLCs, we performed transfection of the reporter constructs during the ESC-to-EpiLC transition. Compared to transfection in serum ESCs passaged in parallel, fragment E remained potent as an enhancer in EpiLCs in conjunction with either the SV40 promoter or endogenous *Tet1* promoter A (fragment A1) (Fig. 2E). These results suggest two functionally autonomous domains within the *Tet1* "superenhancer," each regulating a distinct TSS: (i) the distal 6.2-kb domain B, which appears to be sufficient to activate transcription at exon 1b at a high level, and (ii) the proximal 8.6-kb domain (spanning A and E), comprising a weak promoter at exon 1a and a proximal intragenic enhancer active in both the naive and primed pluripotent states.

A pluripotency-independent promoter coupled to an ESC-specific enhancer in *Tet2*. By 5′ RACE of *Tet2*, we confirmed a dominant TSS at exon 1, 56 kb upstream of the annotated CDS, which aligns with ChIP-seq peaks for Pol2_5P in ESCs (see Fig. S5A and S5B in the supplemental material). To define the functional promoter region, we examined a 6-kb genomic region upstream and inclusive of the 191-bp exon 1 (denoted sequence F) and subcloned 4 fragments serially truncated from the 5′ end (Fig. 3A). All four fragments showed high promoter activity in both ESCs and NIH 3T3 cells (Fig. 3B). Although the promoter activity of each fragment was 2- to 4-fold ($n = 3$ to 5) higher in ESCs than in NIH 3T3 cells, the normalized luciferase activities measured in NIH 3T3 cells were nonetheless substantial (50-fold above those of controls without the promoter), suggesting that the promoter can be functional in either pluripotent or nonpluripotent cells without additional *cis*-regulatory elements. The removal of segment F1 from the 3′ ends of fragments F2 and F3 abolished all promoter activity in ESCs and NIH 3T3 cells (data not shown). These data suggest that the proximal promoter of *Tet2* can be defined within a narrow locus of 556 bp surrounding the TSS, consistent with the high density of active H3K4me3 and H3K27ac ChIP-seq signals in this vicinity and the dearth of these signals upstream (Fig. 3A).

We previously found an O/S motif within a CNS at base position +1788 to +1802 downstream of the *Tet2* TSS; this locus (CNS1) was associated with weak Oct4 binding (13, 25). Two distal intergenic CNS regions that contain putative O/S motifs, at kb −140 (CNS2) and kb −200 (CNS3) upstream (see Fig. S6A in the supplemental material), showed no detectable Oct4 binding but were subsequently described to be functional enhancers during B cell-to-macrophage *trans* differentiation (41). We tested enhancer activity in fragments flanking each O/S motif in the three CNS regions. None of them showed any enhancer activity (<2-fold increase of SV40 promoter activity; $n = 3$) in ESCs (see Fig. S6B in the supplemental material). We also examined a longer 2.5-kb intragenic region, fragment G, that overlaps the motif in CNS1 and the peaks for H3K4me1 and H3K27ac (Fig. 3A). Fragment G was also devoid of enhancer activity in ESCs and NIH 3T3 cells (see Fig. S6C in the supplemental material).

To search for other possible ESC-specific *Tet2* enhancer elements, we looked for ChIP-seq peaks elsewhere in the gene. The superenhancer cluster marked by prominent OSN peaks indeed lies further downstream in introns 1 and 2 (see Fig. S5C in the supplemental material). Based on these OSN locations, we examined 3 regions, regions H, I, and J (Fig. 3A). The 2.8-kb region H showed strong ESC-specific enhancer activity (>10-fold above control levels; $n = 3$); in contrast, regions I and J had weak or no activity (<3-fold; $n = 3$ or 4) (Fig. 3C; see also Fig. S6C in the supplemental material). By splitting region H into two 1.1- to 1.2-kb subfragments, fragments H1 and H2, each corresponding to an individual OSN peak, we further localized the ESC-specific enhancer within fragment H1 (Fig. 3C). Similar enhancer activities (or the lack thereof) were observed for each fragment cloned in either orientation (see Fig. S6D in the supplemental material). When cloned downstream of reporter constructs driven by the already strong promoter fragments F1 and F4, fragment H1 caused a further 3- to 4-fold increase in reporter activities (Fig. 3D). In concert with the loss of *Tet2* expression during the ESC-to-EpiLC transition (see Fig. S2E in the supplemental material), the enhancer activity of fragment H1 was rapidly downregulated in EpiLCs compared to the levels in serum ESCs (Fig. 3E). When enhancer effects on the same mammalian promoter (fragment A5) were compared, *Tet2* enhancer H1 was inactivated, in contrast to the sustained activity of *Tet1* enhancer E in EpiLCs (Fig. 3F). Our results suggest that *Tet2* is regulated by a pluripotency state-independent promoter, but expression can be further increased in ESCs via a distal intragenic naive-state-specific enhancer.

Contribution of Oct and Sox motifs in *Tet1* and *Tet2* enhancer fragments. To determine whether the O/S motif in *Tet1* fragment E (Fig. 4A) binds Oct4 and/or Sox2 *in vitro*, we performed gel shift assays using a 48-bp DNA probe incubated with nuclear extracts from mouse ESCs or from HEK293T cells overexpressing OCT4 and/or SOX2 (Fig. 4B). Two prominent bands of DNA-protein complexes were generated by ESCs and OCT4-expressing HEK293T cell extracts. The lower band was lost and "supershifted" by incubation with an Oct4-specific antibody, which is indicative of a specific DNA-Oct4 complex (Fig. 4B). The higher band (Fig. 4B, asterisk), produced by all nuclear extracts, likely contains the ubiquitously expressed Oct1 (42). Incubation with nuclear extracts of HEK293T cells overexpressing SOX2 failed to produce any additional band that could be supershifted by a Sox2 antibody.

We further mutated the Oct4 motif as well as the adjacent 5′ sequence that weakly resembles a Sox2 motif [C(A/T)TTGT] (Fig. 4A) (28). In competition binding assays, the formation of complexes between the wild-type (wt)-labeled probe and nuclear extracts from ESCs and HEK293T cells expressing OCT4 and SOX2 was effectively abolished by cold wt and Sox motif-mutated (Sm) competitors but not by a cold Oct4 motif-mutated (Om) competitor probe (Fig. 4C; see also Fig. S7A in the supplemental material). We conclude that Oct4 binds directly to its cognate octamer element within fragment E, but *in vitro* Sox2 binding could not be detected under the conditions of our assay.

To test whether the Oct4 motif is functionally critical for enhancer activity, we first divided fragment E into two parts, in which the Oct4 motif was in the 3′ portion. Surprisingly, the 5′ half retained enhancer activity without the Oct4 motif, but the 3′

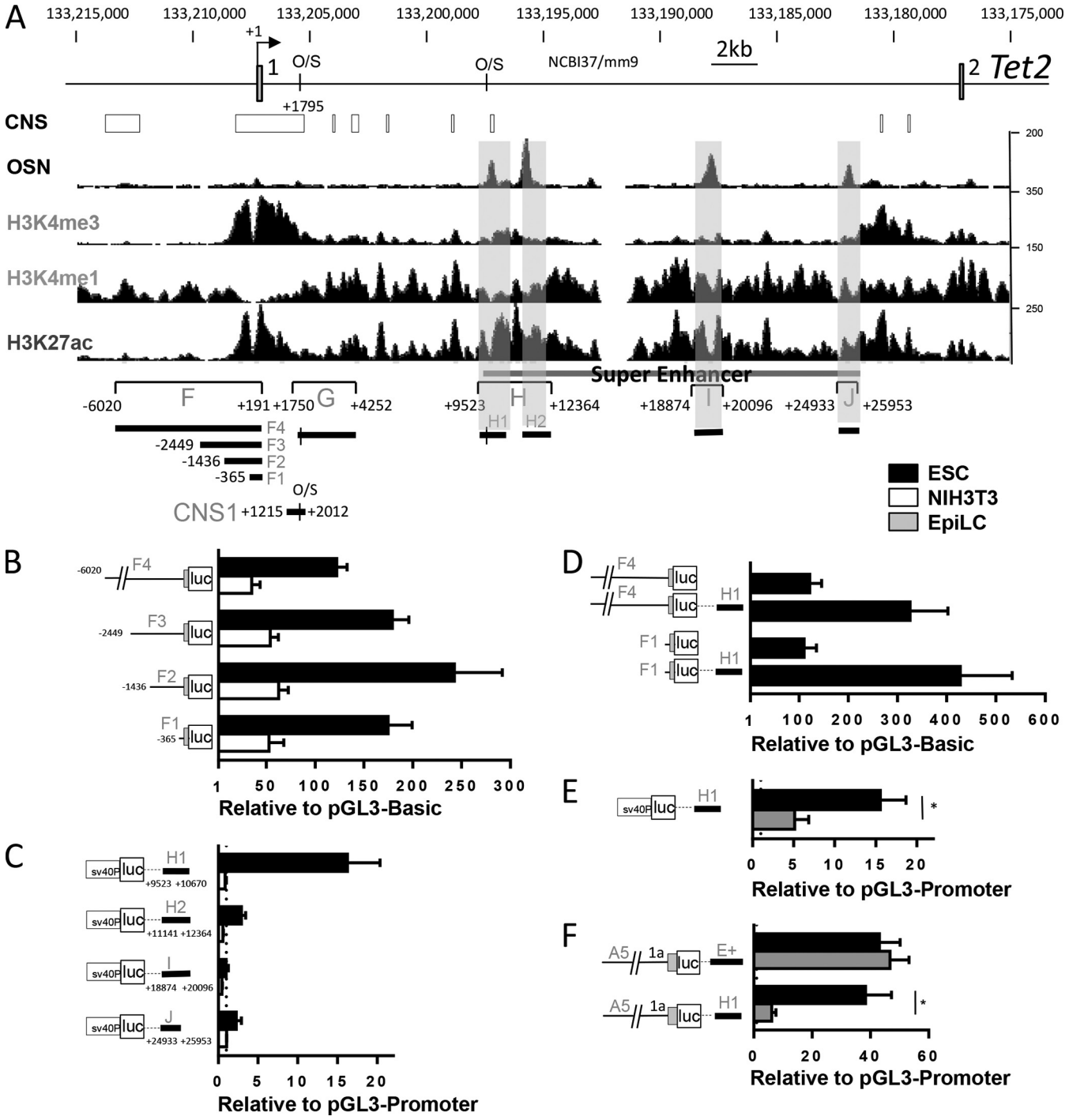
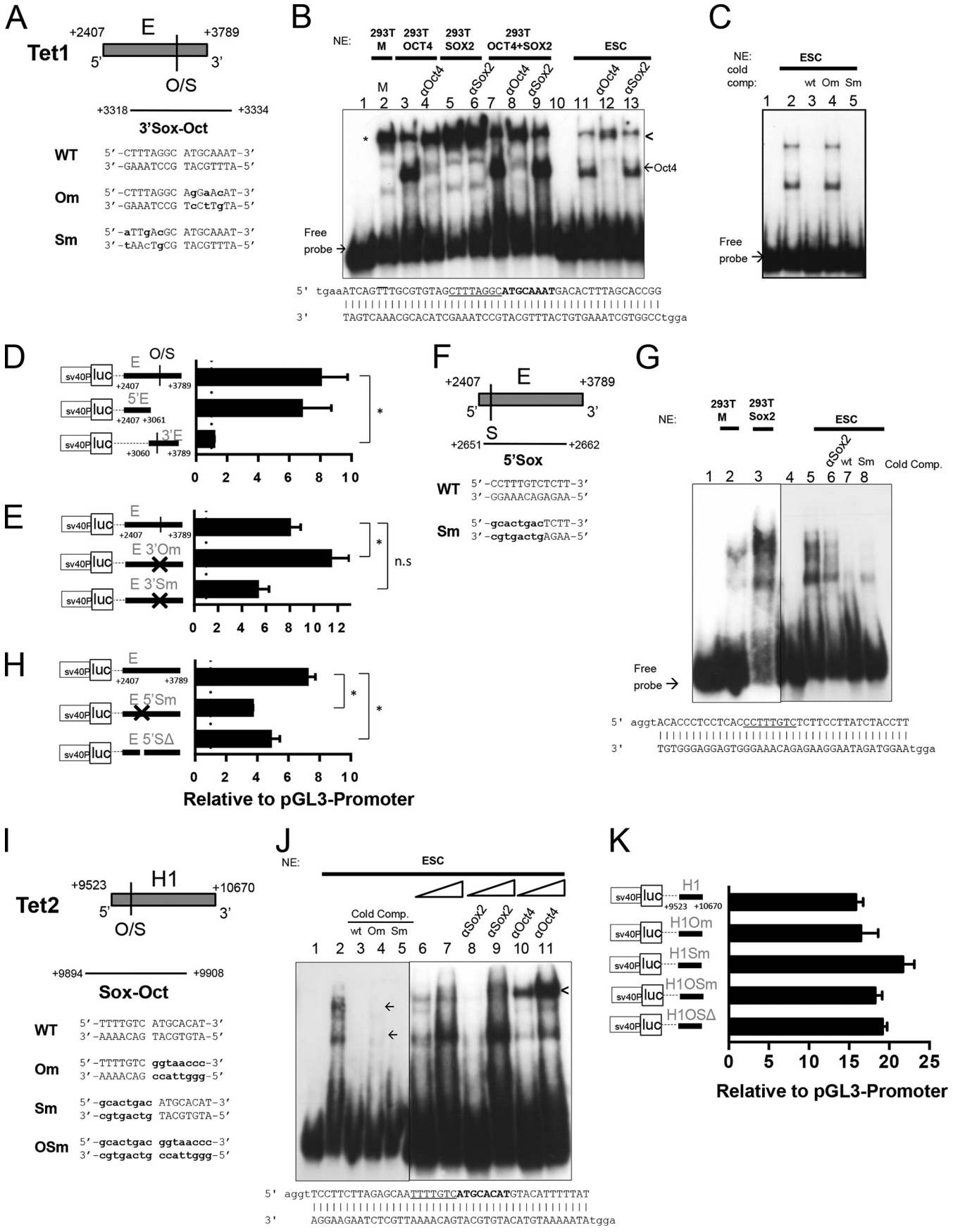


FIG 3 Characterization of *Tet2* regulatory domains by luciferase assays. (A) Same as the legend to Fig. 2A with respect to *Tet2*. (B to F) Normalized luciferase expression of constructs in ESCs, NIH 3T3 cells, and EpiLCs. *, $P < 0.05$.

half lost all stimulatory effects (Fig. 4D). Mutation of the Oct4 motif within the full fragment E also did not abolish but instead increased enhancer activity, whereas mutation of the Sox2 motif had no significant effect ($n = 5$) (Fig. 4E; see also Fig. S7B in the supplemental material). Thus, both deletion and mutation analyses showed that the Oct4 motif in fragment E, despite binding Oct4 *in vitro*, is dispensable for enhancer function.

The lack of enhancer function at the 3′ end of fragment E prompted us to examine the 5′ part for other putative elements. Indeed, we identified a sequence (CCTTTGT) resembling a Sox2 motif (Fig. 4F). Using extracts from ESCs and HEK293T cells overexpressing SOX2, we observed complexes binding to a DNA probe containing the putative motif (labeled 5′ Sox to distinguish from the 3′ Sox-Oct motif shown in Fig. 4A). The Sox2 antibody



caused a variable blockage of complex formation but did not cause any detectable supershift (Fig. 4G; see also Fig. S7C in the supplemental material). Both cold wt and Sox motif mutant probes competed with the labeled probe, although the mutant probe was less effective than the wt, suggesting that Sox2 likely binds *in vitro* to sites other than its putative motif within the probe. Nonetheless, mutation and deletion of the 5′ Sox motif significantly reduced the enhancer activity of fragment E on either the SV40 promoter or the endogenous *Tet1* promoter A (fragment A5) by ~50% (Fig. 4H; see also Fig. S7B in the supplemental material). Together, our results support that Sox2 binds at a different site (5′) from the Oct4 site (3′) within fragment E; however, binding is not fully specific to the Sox motif, consistent with a partial contribution of the motif to enhancer function. In addition to the Sox2 motif, a putative Nanog motif (GNNCATTNNC) (8) was found at the 5′ end of fragment E (see Fig. S7D in the supplemental material). Despite possible Nanog binding to a probe sequence containing the motif, mutation or deletion of the motif had no effect on fragment E enhancer activity (see Fig. S7B, E, and F in the supplemental material).

Tet2 enhancer fragment H1 harbors a putative Oct-Sox motif (Fig. 4I). A probe containing this motif indeed formed complexes with ESC extracts in gel shift assays (Fig. 4J). Incubation with an Oct4 antibody but not a Sox2 antibody produced a supershift. However, both cold Om and Sm probes competed with the labeled probe for binding, although competition by the Om probe appeared to be partial, suggesting that Oct4 binding can also be motif independent. Neither mutation nor deletion of the entire Oct-Sox motif affected fragment H1 enhancer activity *in vitro* (Fig. 4K), suggesting that Oct4 may regulate the *Tet2* enhancer in an motif-independent manner.

Cell type-dependent regulation of *Tet1* by DNA methylation. Murine *Tet1* exon 1a overlaps a CpG island typical of many mammalian promoters and thus is expected to contain the canonical TSS. A previously reported Tet1 ChIP-seq data set also demonstrated Tet1 binding at exon 1a (see Fig. S3 in the supplemental material), suggesting that Tet1 regulates its own promoter (43). In contrast, exon 1b is associated with a distal CpG-poor region with low-level conservation (<50%) with the human genome. In luciferase assays, we noted that B fragments had low but detectable basal activity in NIH 3T3 cells (~10% of ESC levels), which may be attributed to a core promoter activity (Fig. 2B). However, *Tet1* transcripts of exon 1b were not detected in mouse embryonic fibroblasts (MEFs) (Fig. 1D). DNase I hypersensitivity site (DHS) profiles (44) revealed that chromatin accessibility in segment B is very high in ESCs, with individual DHS peaks corresponding to the OSN peaks, but is lost in NIH 3T3 cells (Fig. 5A). Base-resolution whole-genome bisulfite sequencing (WGBS) maps (45) also revealed differential methylation between ESCs and neuronal progenitor (NP) cells within this region, which is indicative of promoter silencing during differentiation (Fig. 5A). We therefore investigated whether there is epigenetic regulation at the promoter regions of *Tet1* by DNA methylation.

We examined the methylation status at CpG sites within or flanking four of the DHS/OSN peaks (D1- to D4-bis denoting loci analyzed) by bisulfite sequencing. All four sites were unmethylated or methylated at low levels (<20%) in ESCs and completely unmethylated in 2i cultures (Fig. 5B and data not shown). Methylation increased marginally in EpiLCs but reached high levels (>70%) in EpiSCs, MEFs, and adult tissues (Fig. 5B; see also Fig. S8A in the supplemental material). In native E6.0 to E7.0 epiblasts, the D1-bis site was already highly methylated, consistent with the diminished levels of exon 1b transcripts compared to those in ESCs and EpiLCs. Traditional bisulfite sequencing cannot discriminate between 5mC and 5hmC among "methylated" cytosines (46). However, global 5hmC levels are very low relative to those of 5mC in murine fibroblasts (13, 38). Moreover, Tet-assisted bisulfite sequencing (47) of selected methylated loci showed hardly any 5hmC (see Fig. S8B in the supplemental material). To confirm that methylation (and not hydroxymethylation) causes silencing of the promoter region B, we performed *in vitro* methylation of fragments B1 to B3, including an additional fragment containing the D1-bis site (fragment B2+D1-bis), in the reporter plasmids by using CpG methyltransferase. Transfection of the methylated plasmids containing all promoter B fragments resulted in markedly reduced levels of reporter activity compared to that in unmethylated plasmids, down to the basal levels driven by the minimal SV40 promoter, which was unaffected by methylation (Fig. 5D). In fact, methylation of only 4 CpG sites in fragments B1 and B2, which are within an OSN ChIP-seq peak and the DHS most proximal to exon 1b (Fig. 2A and 5A), was sufficient to reduce reporter activity by 90%. Together, these results are reflective of a differentially methylated regulatory (DMR) domain within *Tet1* promoter region B that regulates transcription from exon 1b. Consistent with dynamic developmental-stage-specific changes in DNA methylation confined mostly to CpG-poor regulatory regions (48), these results emphasize the need for base resolution and a CpG density-unbiased approach to unravel new critical DMR domains at specific gene loci.

We also examined a region flanking the CpG island at exon 1a

FIG 4 Contribution of Oct-Sox motifs to *Tet1* and *Tet2* enhancer function. (A) The O/S motif in *Tet1* fragment E with mutations within the Oct4 (Om) and Sox2 (Sm) motifs. (B) Electrophoretic mobility shift assay (EMSA) of a *Tet1* probe incubated with nuclear extracts (NE). Lanes: 1 and 10, free probe only; 2, mock-transfected HEK293T cells (M); 3 to 9, HEK293T cells overexpressing either OCT4, SOX2, or OCT4 and SOX2; 11 to 13, v6.5 ESCs. Supershifts were performed by adding Oct4 (lanes 4, 8, and 12)- or Sox2 (lanes 6, 9, and 13)-specific antibodies. The supershifted Oct4-DNA complex is marked with an arrowhead. The topmost band (asterisk) indicates the shift by Oct1. At the bottom, the 48-bp probe sequence is shown with the Oct4 motif in boldface type and the Sox2 motif is underlined. (C) Competition EMSA of the labeled *Tet1* probe (as shown in panel B) incubated with NEs from ESCs (lanes 2 to 5) with the indicated cold competitor oligonucleotides. (D and E) Enhancer activity of *Tet1* fragment E upon truncations (D) or mutations (E) of the O/S motif in ESCs. *, $P < 0.05$ by matched Dunnett's test; n.s., not significant. (F) Sox2 motif at the 5′ end of fragment E and the mutated sequence Sm. (G) EMSA of the labeled *Tet1* 5′-Sox probe (sequence at the bottom with the motif underlined) incubated with NEs from mock-transfected HEK293T cells (lane 2), HEK293T cells overexpressing SOX2 (lane 3), or mouse ESCs (lanes 5 to 8) with the addition of Sox2 antibody or competitor probes, as indicated. (H) Enhancer activity of *Tet1* fragment E upon mutation or deletion (Δ) of the 5′-Sox2 motif in ESCs. *, $P < 0.05$ by matched Dunnett's test. (I) Putative O/S motif in *Tet2* fragment H1 with mutations within the Oct4 (Om) and Sox2 (Sm) motifs. (J) EMSA of the labeled *Tet2* probe (sequence at the bottom with the Oct4 motif in boldface type and the Sox2 motif underlined) incubated with NEs from mouse ESCs with the addition of antibodies or competitor probes, as indicated. In lanes 6 to 11, NEs were loaded in two amounts (2 or 4 μl), as indicated by the rising bars. The supershifted Oct4-DNA complex is marked with an arrowhead. The arrows (lane 4) mark the bands partially competed by the cold Om probe. (K) Enhancer activity of *Tet2* fragment H1 upon mutation or deletion (Δ) of the O/S motif in ESCs.

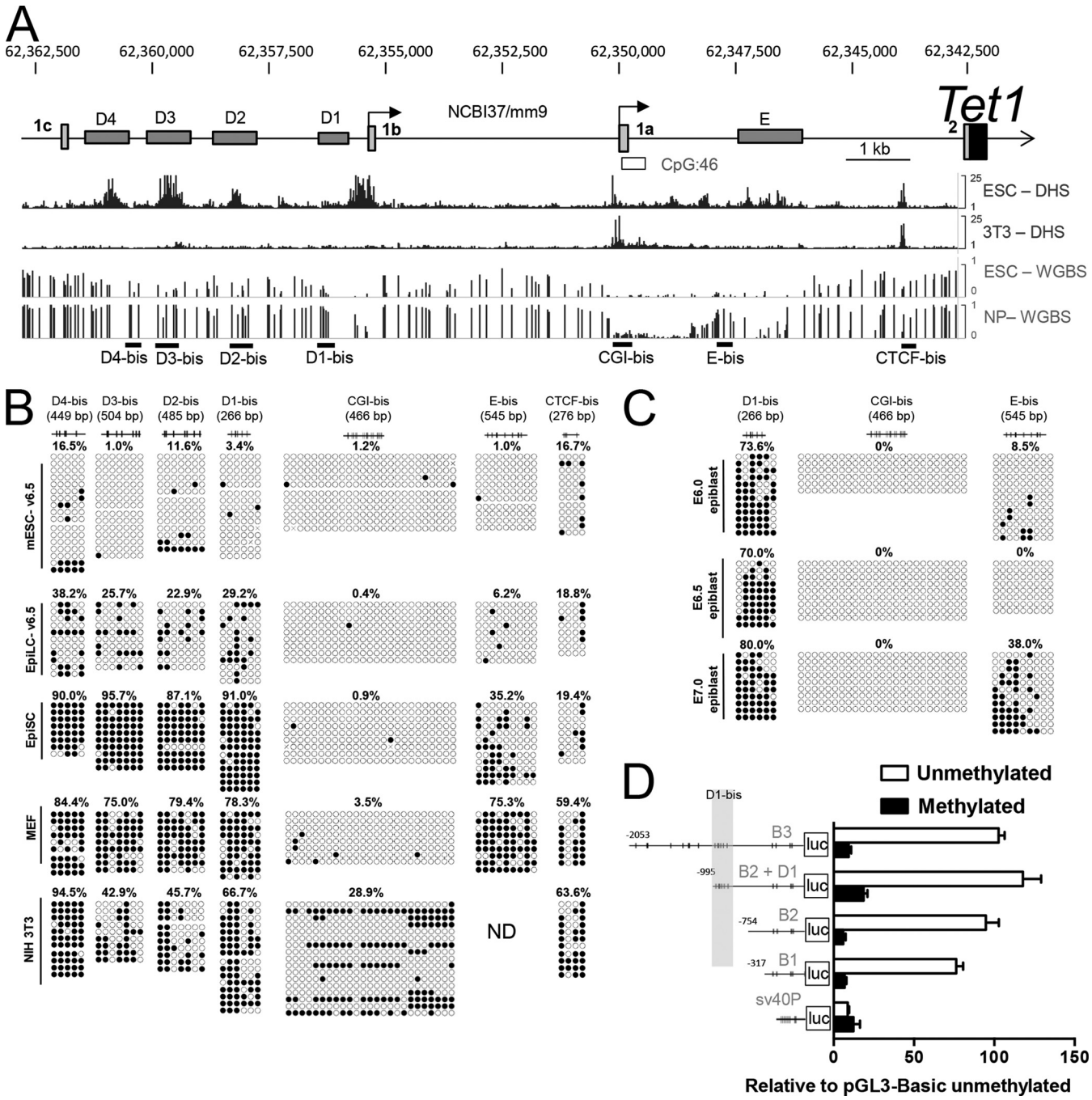


FIG 5 Analysis of DNA methylation at *Tet1* sites by bisulfite sequencing. (A) The 5′ genomic region of *Tet1* (Fig. 2A) showing domains of interest. The white box indicates the position of a CpG island containing 46 CpG sites. Below the scheme are ENCODE tracks for DHSs (44) and WGBS methylome maps (45) of mouse ESCs versus NIH 3T3 cells or neural progenitors (NP). At the bottom, loci for bisulfite analysis are indicated by horizontal lines. (B and C) Methylation of specific *Tet1* loci (size shown in base pairs in parentheses and with relative positions of CpGs marked by vertical lines) in v6.5 ESCs, EpiLCs, EpiSCs, MEFs, and NIH 3T3 cells (B) and in E6.0 to E7.0 epiblasts (C). Closed circles denote 5mC (or 5hmC); open circles indicate unmethylated C; × indicates a missing sequence or mismatch. Results are representative of 2 or 3 independent experiments. Where there are <7 clonal sequences from each conversion, sequences from additional experiments (separated by a space) are added. Values at the top indicate overall methylation levels (percent) of the sequences shown. ND, not determined. (D) Effect of DNA methylation on normalized luciferase activities of reporter plasmids.

of *Tet1* (CGI-bis). Clonal bisulfite analysis of 25 of the 46 CpG sites within the CpG island revealed a similar low-methylation state (<15%) in all embryonically derived and somatic cells, including adult liver and brain tissues (Fig. 5B and C; see also Fig. S8A in the supplemental material). In NIH 3T3 cells, the levels of methylation were higher in region A (28.9%) but lower in region B (represented by D1- to D4-bis) than those detected in MEFs, which is suggestive of aberrant *de novo* methylation at the CpG

island and hypomethylation at CpG-poor sites unique to tumor transformation. DHS profiles also showed that this is one of few regions at the *Tet1* locus, including one site ~6 kb downstream with a differential CCCTC binding factor (CTCF) binding pattern, that retains an open chromatin in both ESCs and NIH 3T3 cells (Fig. 5A). An adjacent locus 5′ of the CpG-poor enhancer E was methylated at low levels (<20%) in ESCs, EpiLCs, and E6.0 to E6.5 epiblasts; it gained intermediate methylation states (20 to 40%) in EpiSCs and E7.0 epiblasts and became highly methylated in MEFs (75.3%) and other somatic tissues (liver [94.4%] and brain [87.7%]) analyzed (Fig. 5B and C; see Fig. S8A in the supplemental material). Similarly, the CTCF binding site further downstream was methylated at low levels (<20%) in ESCs, EpiLCs, and EpiSCs but at high levels in MEFs (59.4%) and NIH 3T3 cells (63.6%), in an inverse relation to CTCF binding (Fig. 5B and data not shown). Thus, the entire gene locus upstream of *Tet1*, except for the CpG island region, eventually gained methylation associated with the silencing of promoter region B early and enhancer region E later during differentiation.

The active TSS of *Tet2* is flanked by a high-density CpG island. Both DHS and WGBS profiles for this genomic region suggest an open chromatin with low-level methylation in both pluripotent and differentiated cells (see Fig. S8C in the supplemental material). Clonal bisulfite sequencing analysis of 33 of 115 CpG sites within the CpG island indeed revealed that this locus was completely unmethylated in ESCs, EpiLCs, EpiSCs, and MEFs, although there was a notable increase in methylation (to 37.9%) in NIH 3T3 cells (see Fig. S8D in the supplemental material). The intragenic enhancer region H1 was methylated at low levels in ESCs, EpiLCs, and E6.0 to E7.0 epiblasts (<15%) but was methylated at intermediate levels (30 to 50%) in EpiSCs and MEFs (see Fig. S8D and S8E in the supplemental material), suggesting that this enhancer may not be regulated by DNA methylation during exit from pluripotency.

We further investigated the dynamics of transcriptional and methylation changes at *Tet1* during differentiation using an *in vitro* assay of LIF and feeder withdrawal and supplementation with all-*trans* retinoic acid (RA). As previously reported, *Tet1* and *Tet2* transcripts were rapidly downregulated concomitantly with *Oct4* within 2 days of LIF withdrawal and RA induction (13). We measured *Tet1* transcript levels from exons 1a and 1b and coding regions during this time course by qPCR. Exogenous LIF sustained *Tet1* expression for at least 3 to 4 days, although total transcript levels progressively decreased to 2 to 4% of *Gapdh* levels upon feeder-free adaptation (Fig. 6A). Upon RA-induced differentiation, the expression of exon 1b rapidly diminished within 24 h and was undetectable by 2 days. In contrast, the expression of exon 1a persisted at lower levels in both LIF-cultured and differentiating cells (Fig. 6A). Thus, the decline of total *Tet1* levels during RA-induced differentiation is reflected by changes in exon 1b and not exon 1a transcripts.

Consistent with sustainable exon 1a expression, the CpG island region proximal to exon 1a stayed unmethylated during the entire differentiation time course (see Fig. S9A in the supplemental material). The downstream CTCF site also stayed methylated at similarly low levels (<40%) in both the LIF-sustained and RA-treated (in the absence of LIF) groups (see Fig. S9B in the supplemental material). Although the CTCF site stayed unmethylated during this early differentiation time course, it was methylated later in the somatic lineage (Fig. 5B).

To track the methylation changes in genomic region B of *Tet1*, we focused on the two DMR sites D1- and D4-bis. In the presence of exogenous LIF, methylation levels at site D1 stayed low (<30%), at a level typical of feeder-free adapted ESCs grown in serum, and increased only marginally to 34.8% at day 3, along with increased spontaneous differentiation. Upon RA-induced differentiation, methylation exceeded 50% by 48 h and reached a high level of 70% by day 3 (Fig. 6B). Overall, the kinetics of methylation gain at site D1-bis appeared to be tightly associated with the loss of exon 1b expression and further supports DNA methylation at this locus as a transcriptional silencing mark. In contrast, methylation of the D4 locus showed fewer differential changes during differentiation (Fig. 6C). Under both LIF-supplemented and RA-induced differentiating conditions, site D4-bis progressively increased methylation to 50 to 60% during the first 2 days and displayed differential methylation only at day 3 (48.0% versus 80.0%). Thus, the methylation changes at site D4 did not correlate with transcript levels of exon 1b but may prime the region for silencing by differentiation cues. From these kinetics analyses of methylation changes at selected loci in *Tet1*, we infer that a wave of chromatin compaction progresses from the distal border toward the gene body during cellular differentiation but preserving a constitutive open region in the CpG island at exon 1a.

Sufficiency of *Tet1* promoter domain B. To determine whether the 6-kb genomic segment B of *Tet1* can function autonomously to mimic physiological patterns of gene expression, we placed fragment B6 upstream of the mCherry fluorescence reporter and stably introduced the construct into ESCs (Fig. 7A). Two puromycin-resistant clones with bright fluorescence were selected for further analysis. Flow cytometry showed that >80% of cells from feeder cultures expressed the reporter, although all of them expressed high levels of the naive surface marker CD31 (Fig. 7B and data not shown; see also Fig. S10 in the supplemental material). We speculate that ~20% of the cells from each clone may have undergone epigenetic silencing at the transgene during culture, even in the absence of spontaneous differentiation. When these clones were feeder depleted and induced to differentiate by using the same RA treatment as the one described above, fluorescence diminished in 2 days, similar to the kinetics of downregulation of the endogenous exon 1b transcripts (Fig. 7C and D; see also Fig. S10 in the supplemental material). Thus, our *Tet1* fragment B6 transgenic reporter system faithfully mimics endogenous promoter activity, at least *in vitro*, and shows that the 6.2-kb promoter fragment B is sufficient to regulate transcription from exon 1b.

DISCUSSION

In this study, we defined two distinct TSSs in *Tet1* and additional *cis*-regulatory elements in both *Tet1* and *Tet2* that account for their tissue- and development-stage expression patterns. The dynamic coupling of active promoter and enhancer domains in *Tet1* causes a switch in dominant TSS usage during the transition from naive to primed pluripotent states, when expression from the gene-proximal TSS is sustained by a neighboring enhancer, and a further loss of expression in differentiated cells when both distal promoter and proximal enhancer regions are silenced. In contrast, *Tet2* transcription is driven by a non-ESC-specific promoter, but high levels are sustained in naive ESCs by a distal intragenic enhancer that is rapidly inactivated in primed epiblast-like cells.

Our results account for the different tissue-specific expression patterns of *Tet1* and *Tet2* previously reported in the mouse (37, 38,

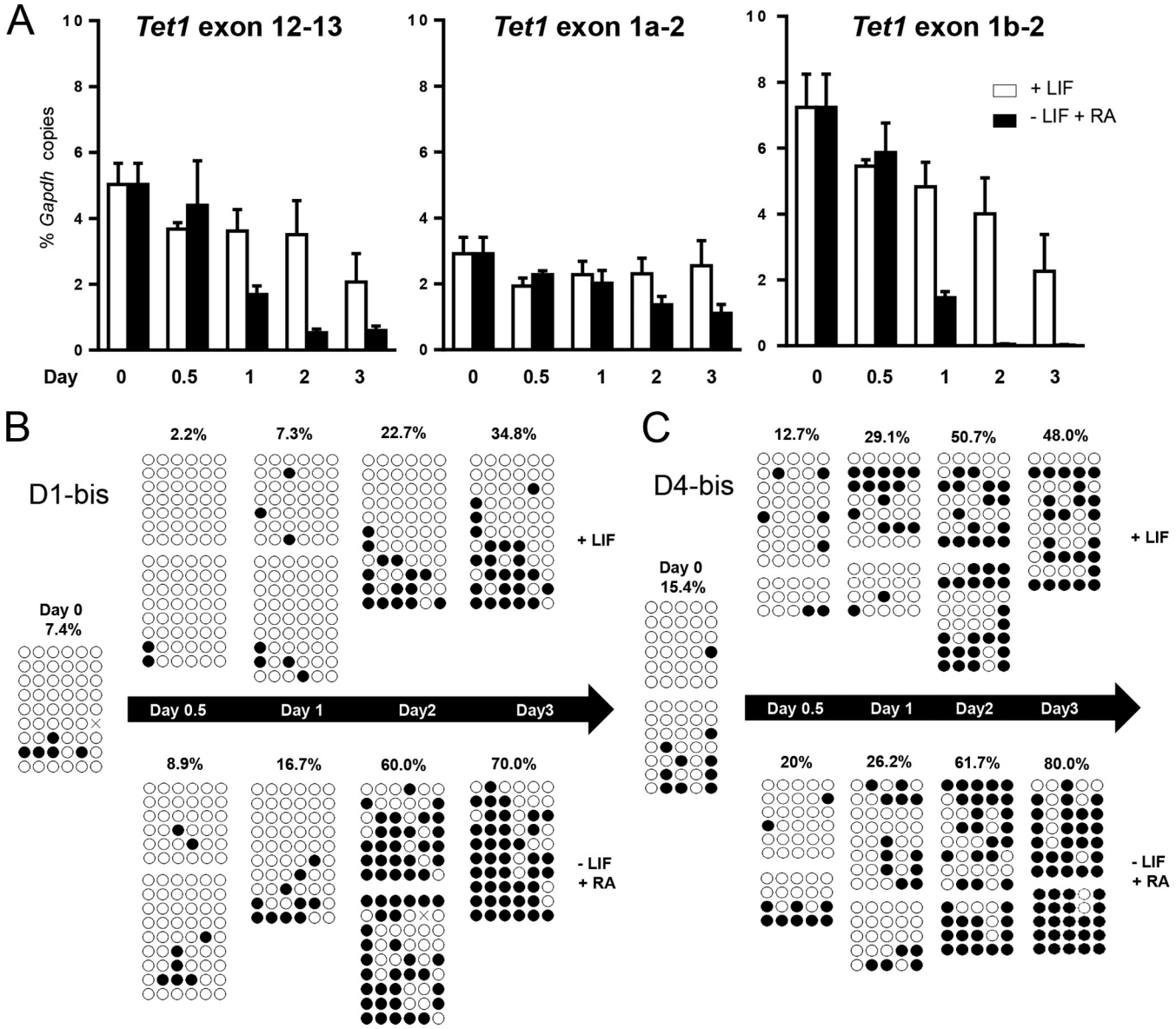


FIG 6 Time course of DNA methylation changes at *Tet1* sites during *in vitro* differentiation. (A) Normalized copy numbers of *Tet1* transcripts in ESCs upon LIF withdrawal and the addition of 1 μM retinoic acid (RA). Expression values are means ± SEM ($n = 3$). (B and C) Methylation of two sites (sites D1- and D4-bis) during the time course. Data are representative of 2 independent experiments. Where there are <9 clones, sequences from the second experiment are added.

49). Unlike in the mouse, *TET1* levels remained stable during differentiation of human ESCs (14). Our study resolves this discrepancy by demonstrating that postimplantation mouse epiblast and epiblast-derived stem cells, which share characteristics more similar to those of human ESCs (50), already lost the bulk of *Tet1* expression via exon 1b and sustained *Tet1* expression predominantly by using TSSs at exon 1a. Indeed, 5′ RACE analysis of the H9 human ESC line confirmed that *TET1* transcription starts from the annotated exon 1 homologous to exon 1a in the mouse (data not shown). Thus, the high *Tet1* expression levels observed for mouse ESCs and iPSCs arise from the activation of a second promoter in the naive pluripotent state. Although not tested, we predict that high expression levels of *Tet1* in primordial germ cells (15) are induced via strong transcription initiation from exon 1b. A recent study defined conditions to "reset" human ESCs to ground-state pluripotency, in which elevated *TET1* expression was observed (51). Whether an analogous alternative TSS and promoter region can be activated in naive-state human ESCs remains to be confirmed, but if indeed proven so, they will be valuable tools to discern truly naive-state human cells derived under alternative culture conditions.

Recent methylome analyses revealed that mouse 2i ESCs develop a global hypomethylated state, associated with the suppression of *Dnmt3a/Dnmt3b/Dnmt3L*, which mimics the early ICM; in comparison, serum-cultured ESCs acquire *de novo* methylation and increased expression of *Dnmt3b* more typical of the epiblast in late blastocysts (32, 33). Those studies analyzed feeder-free ESCs, and it is unclear how feeder-cultured ESCs correlate with the ICM. We consistently observed higher levels of *Tet1* expression from both TSSs in feeder-cultured than in feeder-free ESCs. A previ-

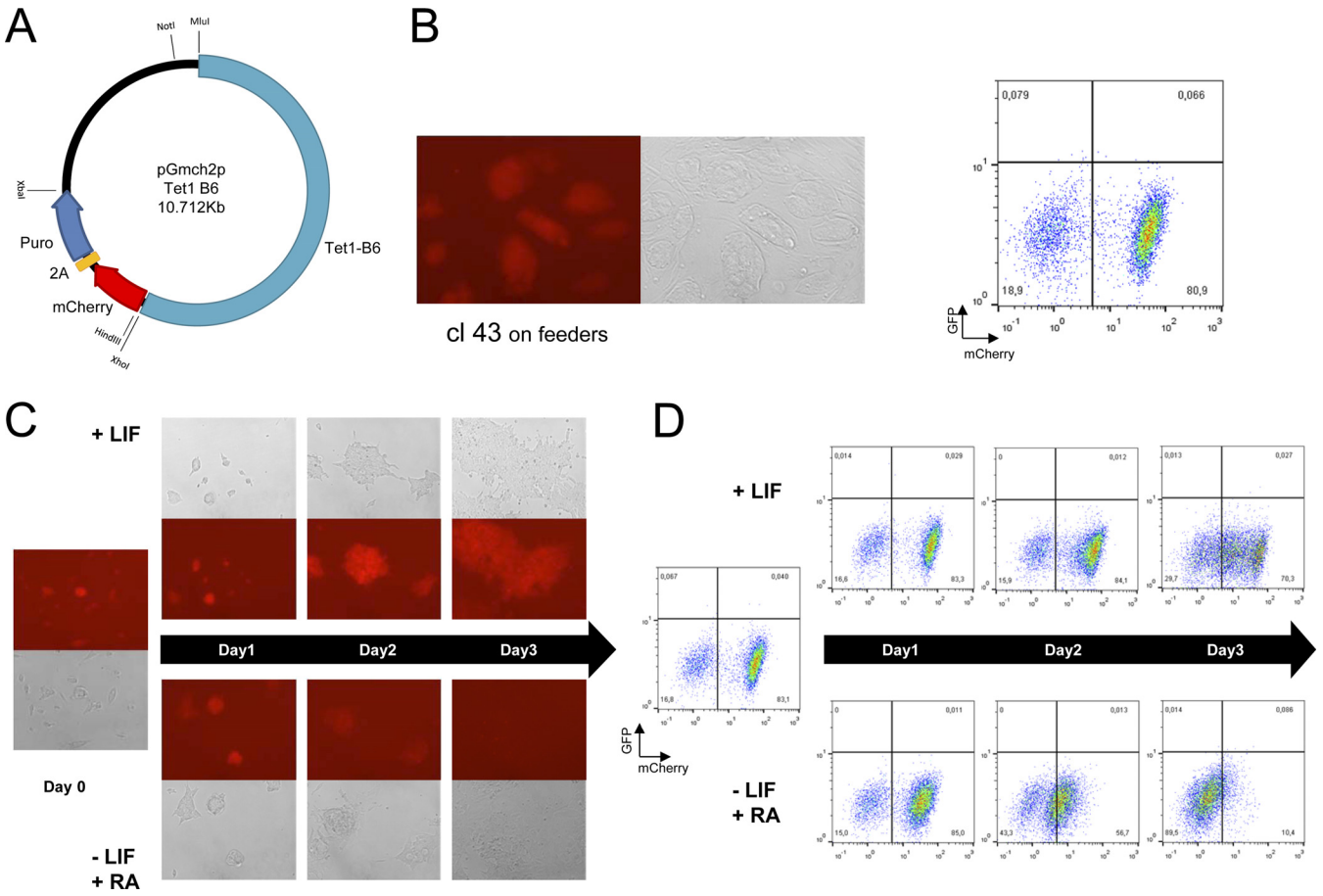


FIG 7 Characterization of a *Tet1* fragment B6 transgene reporter ESC line. (A) Plasmid construct of a 6-kb exon 1b-associated promoter fragment (fragment B6) subcloned to drive expression of mCherry and a drug-selectable marker. (B, left) Fluorescence image (shown next to the bright-field image) of a selected clone grown on feeders. (Right) Flow cytometry. (C and D) Fluorescence images (C) and flow cytometry (D) of cells harvested daily during *in vitro* differentiation. Data are representative of 3 independent experiments.

ously reported single-cell RNA sequencing analysis showed that ESCs (feeder cultured) are molecularly distinct from the E3.5 ICM and E4.5 epiblasts (52) and, by extrapolation, may also be distinct from feeder-depleted ESCs. Nonetheless, our results suggest that preimplantation development in the mouse involves higher *Tet1* transcription levels, which are suppressed by the postimplantation stage. The transition from the preimplantation ICM (E3.5) to postimplantation epiblasts (E6.5) is associated with a global gain of DNA methylation (53). The physiological function of *Tet1* has been proposed to be that of a "maintenance" DNA demethylase to maintain the fidelity of DNA methylation patterns (54, 55). Collectively, our results are consistent with a developmental model in which a burst of *Tet1* transcription accompanies the upregulation of *Dnmt3a/Dnmt3b* in the transition from the early to late ICM to regulate *de novo* DNA methylation.

An epigenome comparison of naive ESCs and primed EpiSCs revealed the presence of "seed" enhancers, which are dormant in ESCs but activated in EpiSCs to take over primary transcriptional control of genes that are similarly expressed in either cell state (7). By this definition, the proximal enhancer E of *Tet1* is neither a seed nor a "poised" enhancer because total *Tet1* RNA levels were much higher in ESCs than in EpiSCs; moreover, enhancer E demonstrated *in vitro* activity in both cell states (Fig. 2E). Nonetheless, a striking analogy to the master pluripotency gene *Oct4* can be drawn. The distal enhancer of *Oct4*, also the densest binding locus for key pluripotency transcription factors in ESCs (56), drives *Oct4* expression in preimplantation embryos, primordial germ cells, and ESCs but not in postimplantation epiblasts, whereas the proximal enhancer directs postimplantation epiblast-specific expression (57).

Enhancers are classically described as position and orientation independent. *Tet1* fragment E is unique in showing significantly higher enhancer activity when cloned in the sense than in the antisense direction *in vitro*. This is inconsistent with the model of transcription factor binding and activation in which recognition of cognate motifs can occur in either orientation (3). We speculate that there is an involvement of enhancer-associated noncoding RNA (ncRNA or eRNA), which is detectable by genomic run-on sequencing (GRO-seq) over the *Tet1* superenhancer (see Fig. S3 in the supplemental material). The precise biological functions of these enhancer-like RNAs, which are transcribed bidirectionally, are only beginning to be understood, including a role in facilitating chromatin looping and transcription via an interaction with Mediator (58). The presence of an SV40 late poly(A) signal in our luciferase construct in proximity to the enhancer cloning site may have interfered with ncRNA production in one direction.

A previous study reported that *Tet2* is regulated by Oct4 (25). The authors of that study cloned a 2.4-kb genomic fragment, from base position −420 upstream of the TSS to base position +1980 downstream in the intronic region encompassing the conserved O/S element in CNS1 (Tet2L) upstream of a reporter. By cotransfecting the reporter plasmid into NIH 3T3 cells with expression plasmids, those authors demonstrated that Oct4, but not Sox2, is necessary for *Tet2* transcription (25). In contrast, we showed that the minimal promoter region of *Tet2* placed immediately upstream of the reporter gene is active in both pluripotent and nonpluripotent cells, even without the octamer element. Our results are consistent with the broad expression of *Tet2* across tissue types in the soma by virtue of its pluripotency-independent promoter.

A previous analysis of ChIP-seq data from mouse ESCs (56) showed the binding of multiple pluripotency-associated transcription factors to a single cluster 12.5 to 22.3 kb upstream of the *Tet1* gene, corresponding to genomic region B defined in this study (59). However, an Oct4 site was lacking in this cluster. Here we used a more recent ChIP-seq data set (8), which revealed similar binding sites within region B for many of these factors, including Oct4. Sequence analysis of the superenhancer region revealed only one O/S composite element in *Tet1* region E. Here we validated that the composite motif indeed binds Oct4 (but not Sox2) directly in gel shift assays, consistent with the 100% match with the Oct4 consensus motif but a much weaker match for Sox2. Surprisingly, both composite Oct4-Sox2 motifs in *Tet1* (fragment E) and *Tet2* (fragment H1) enhancer fragments are not necessary for enhancer function. The lack of contribution by a motif in an *in vitro* luciferase reporter assay does not preclude a function *in vivo* in the native chromatin context. Nonetheless, our results support motif (sequence)-independent binding of transcription factors at high-occupancy target (HOT) regions (60, 61).

By switching *in vitro* culture conditions of ESCs to mimic different states of embryonic development, together with analyses of early-postimplantation-stage epiblasts, we propose that the transition from pre- to postimplantation development involves a dramatic change in the transcriptional landscape at both the *Tet1* and *Tet2* gene loci so as to sustain *Tet1* expression via TSSs at exon 1a but rapidly downregulate *Tet2* in the postimplantation, prestreak epiblast. In view of the very low levels of *Tet2* and *Tet3* expression in the pre- and early-streak mouse epiblasts, *Tet1* appears to be the predominant 5-methylcytosine oxygenase expressed at these stages (R. Khoueiry and K. P. Koh, unpublished data). In contrast to the pluripotency genes *Oct4* and *Nanog*, *Tet1* levels may decline rapidly during differentiation but are not completely silenced. The presence of a non-cell type-selective, albeit weak, promoter at exon 1a would contribute to low-level Oct4-independent expression of *Tet1* in the postgastrulation embryo and adult soma, particularly in the brain (62, 63). Full functional validation of the *cis*-regulatory elements described in this study will require further studies to remove specific sequences in cells and in transgenic mice. The precise physiological relevance of *Tet1* expression in the early embryo also warrants further investigation by using strategies to abolish gene expression completely in the mouse.

Recent studies support a role of *Tet1* and *Tet2* in epigenetic reprogramming by facilitating DNA demethylation of *Oct4* and other genes; Tet1 can even replace Oct4 in the classical 4-factor cocktail (64). Many studies have relied on Oct4-green fluorescent protein (GFP) and Nanog-GFP reporters to select iPSCs. New transgenic reporters based on *Tet1* and *Tet2* will be valuable tools to unravel the mechanisms of their gene activation during reprogramming and ultimately provide insights into the acquisition of different induced pluripotent states and epigenetic memory.

ACKNOWLEDGMENTS

We thank Catherine Verfaillie (KU Leuven) and Alice Jouneau (INRA, France) for generously providing reagents and critical reading of the manuscript, Elke Stappers and Danny Huylebroeck (KU Leuven) for guidance in dissection of mouse brain tissues, and Aernout Luttun (KU Leuven) for the use of a luminometer.

Research in the laboratory of Kian Peng Koh is supported by the Fonds voor Wetenschappelijk Onderzoek (FWO) Research Foundation—Flanders (grants G.0C56.13N and G.0632.13), the Ministerie van de Vlaamse Gemeenschap, and a Marie Curie career integration grant (PCIG-GA-2012-321658). Frank Classens and Lien Spans are supported by the FWO (grant G.0684.12N). Sridhar Rao is funded in part by the Midwest Alliance against Childhood Cancer, Hyundai Hope on Wheels, the Children's Research Institute, and an Institutional Research Grant (86-004) from the American Cancer Society.

REFERENCES

1. **Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, et al.** 2012. Landscape of transcription in human cells. Nature **489:**101–108. http://dx.doi.org/10.1038/nature11233.
2. **Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B.** 2012. A map of the cis-regulatory sequences in the mouse genome. Nature **488:**116–120. http://dx.doi.org/10.1038/nature11243.
3. **Spitz F, Furlong EEM.** 2012. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet **13:**613–626. http://dx.doi.org/10.1038/nrg3207.
4. **Kim J, Chu J, Shen X, Wang J, Orkin SH.** 2008. An extended transcriptional network for pluripotency of embryonic stem cells. Cell **132:**1049–1061. http://dx.doi.org/10.1016/j.cell.2008.02.039.
5. **Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T.** 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell **10:**105–116. http://dx.doi.org/10.1016/j.devcel.2005.10.017.
6. **Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, Wysocka J.** 2014. Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell **14:**838–853. http://dx.doi.org/10.1016/j.stem.2014.04.003.
7. **Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, McKay RD, Crawford GE, Scacheri PC, Tesar PJ.** 2014. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell Stem Cell **14:**854–863. http://dx.doi.org/10.1016/j.stem.2014.05.005.
8. **Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA.** 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell **153:**307–319. http://dx.doi.org/10.1016/j.cell.2013.03.035.
9. **He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song C-X, Zhang K, He C, Xu G-L.** 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science **333:**1303–1307. http://dx.doi.org/10.1126/science.1210944.
10. **Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y.** 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science **333:**1300–1303. http://dx.doi.org/10.1126/science.1210597.
11. **Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A.** 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by

MLL partner TET1. Science **324:**930–935. http://dx.doi.org/10.1126/science.1170116.
12. **Reik W.** 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature **447:**425–432. http://dx.doi.org/10.1038/nature05918.
13. **Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A.** 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell **8:**200–213. http://dx.doi.org/10.1016/j.stem.2011.01.008.
14. **Langlois T, da Costa Reis Monte-Mor B, Lenglet G, Droin N, Marty C, Le Couedic JP, Almire C, Auger N, Mercher T, Delhommeau F, Christensen J, Helin K, Debili N, Fuks F, Bernard OA, Solary E, Vainchenker W, Plo I.** 2014. TET2 deficiency inhibits mesoderm and hematopoietic differentiation in human embryonic stem cells. Stem Cells **32:**2084–2097. http://dx.doi.org/10.1002/stem.1718.
15. **Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA.** 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science **339:**448–452. http://dx.doi.org/10.1126/science.1229277.
16. **Vincent JJ, Huang Y, Chen PY, Feng S, Calvopina JH, Nee K, Lee SA, Le T, Yoon AJ, Faull K, Fan G, Rao A, Jacobsen SE, Pellegrini M, Clark AT.** 2013. Stage-specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell **12:**470–478. http://dx.doi.org/10.1016/j.stem.2013.01.016.
17. **Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, Gao Q, Powell BE, Li Z, Xu M, Faull KF, Lyko F, Jaenisch R.** 2013. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell **24:**310–323. http://dx.doi.org/10.1016/j.devcel.2012.12.015.
18. **Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y.** 2013. Role of Tet1 in erasure of genomic imprinting. Nature **504:**460–464. http://dx.doi.org/10.1038/nature12805.
19. **Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, Fidalgo M, Saunders A, Lawrence M, Dietmann S, Das S, Levasseur DN, Li Z, Xu M, Reik W, Silva JC, Wang J.** 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature **495:**370–374. http://dx.doi.org/10.1038/nature11925.
20. **Hu X, Zhang L, Mao SQ, Li Z, Chen J, Zhang RR, Wu HP, Gao J, Guo F, Liu W, Xu GF, Dai HQ, Shi YG, Li X, Hu B, Tang F, Pei D, Xu GL.** 2014. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell **14:**512–522. http://dx.doi.org/10.1016/j.stem.2014.01.001.
21. **Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR.** 2003. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia **17:**637–641. http://dx.doi.org/10.1038/sj.leu.2402834.
22. **Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabo PE, Pfeifer GP, Li J, Xu GL.** 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature **477:**606–610. http://dx.doi.org/10.1038/nature10443.
23. **Hsu C-H, Peng K-L, Kang M-L, Chen Y-R, Yang Y-C, Tsai C-H, Chu C-S, Jeng Y-M, Chen Y-T, Lin F-M, Huang H-D, Lu Y-Y, Teng Y-C, Lin S-T, Lin R-K, Tang F-M, Lee S-B, Hsu HM, Yu J-C, Hsiao P-W, Juan L-J.** 2012. TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. Cell Rep **2:**568–579. http://dx.doi.org/10.1016/j.celrep.2012.08.030.
24. **Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q, Lee CW, Hu D, Lian BQ, Kleffel S, Yang Y, Neiswender J, Khorasani AJ, Fang R, Lezcano C, Duncan LM, Scolyer RA, Thompson JF, Kakavand H, Houvras Y, Zon LI, Mihm MC, Jr, Kaiser UB, Schatton T, Woda BA, Murphy GF, Shi YG.** 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell **150:**1135–1146. http://dx.doi.org/10.1016/j.cell.2012.07.033.
25. **Wu Y, Guo Z, Liu Y, Tang B, Wang Y, Yang L, Du J, Zhang Y.** 2013. Oct4 and the small molecule inhibitor, SC1, regulates Tet2 expression in mouse embryonic stem cells. Mol Biol Rep **40:**2897–2906. http://dx.doi.org/10.1007/s11033-012-2305-5.
26. **Maruotti J, Dai XP, Brochard V, Jouneau L, Liu J, Bonnet-Garnier A, Jammes H, Vallier L, Brons IG, Pedersen R, Renard JP, Zhou Q, Jouneau A.** 2010. Nuclear transfer-derived epiblast stem cells are transcriptionally and epigenetically distinguishable from their fertilized-derived counterparts. Stem Cells **28:**743–752. http://dx.doi.org/10.1002/stem.400.
27. **Pulakanti K, Pinello L, Stelloh C, Blinka S, Allred J, Milanovich S, Kiblawi S, Peterson J, Wang A, Yuan GC, Rao S.** 2013. Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. Epigenetics **8:**1303–1320. http://dx.doi.org/10.4161/epi.26597.
28. **Chew JL, Loh YH, Zhang W, Chen X, Tam WL, Yeap LS, Li P, Ang YS, Lim B, Robson P, Ng HH.** 2005. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol **25:**6031–6046. http://dx.doi.org/10.1128/MCB.25.14.6031-6046.2005.
29. **Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA.** 2010. c-Myc regulates transcriptional pause release. Cell **141:**432–445. http://dx.doi.org/10.1016/j.cell.2010.03.030.
30. **Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA.** 2007. A chromatin landmark and transcription initiation at most promoters in human cells. Cell **130:**77–88. http://dx.doi.org/10.1016/j.cell.2007.05.042.
31. **Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A.** 2008. The ground state of embryonic stem cell self-renewal. Nature **453:**519–523. http://dx.doi.org/10.1038/nature06968.
32. **Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul Y-L, Walter J, Cook SJ, Andrews S, Branco MR, Reik W.** 2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell **13:**351–359. http://dx.doi.org/10.1016/j.stem.2013.06.004.
33. **Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HH, Matarese F, Lepikhov K, Gut M, Brun-Heath I, Hubner NC, Benedetti R, Altucci L, Jansen JH, Walter J, Gut IG, Marks H, Stunnenberg HG.** 2013. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell **13:**360–369. http://dx.doi.org/10.1016/j.stem.2013.06.002.
34. **Hackett JA, Dietmann S, Murakami K, Down TA, Leitch HG, Surani MA.** 2013. Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. Stem Cell Rep **1:**518–531. http://dx.doi.org/10.1016/j.stemcr.2013.11.010.
35. **Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M.** 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell **146:**519–532. http://dx.doi.org/10.1016/j.cell.2011.06.052.
36. **Kojima Y, Kaufman-Francis K, Studdert JB, Steiner KA, Power MD, Loebel DA, Jones V, Hor A, de Alencastro G, Logan GJ, Teber ET, Tam OH, Stutz MD, Alexander IE, Pickett HA, Tam PP.** 2014. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell **14:**107–120. http://dx.doi.org/10.1016/j.stem.2013.09.014.
37. **Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y.** 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature **466:**1129–1133. http://dx.doi.org/10.1038/nature09303.
38. **Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H.** 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res **38:**e181. http://dx.doi.org/10.1093/nar/gkq684.
39. **Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R.** 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A **107:**21931–21936. http://dx.doi.org/10.1073/pnas.1016071107.
40. **Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J.** 2011. A unique chromatin signature uncovers early developmental enhancers in humans. Nature **470:**279–283. http://dx.doi.org/10.1038/nature09692.
41. **Kallin EM, Rodriguez-Ubreva J, Christensen J, Cimmino L, Aifantis I, Helin K, Ballestar E, Graf T.** 2012. Tet2 facilitates the derepression of myeloid target genes during CEBPalpha-induced transdifferentiation of pre-B cells. Mol Cell **48:**266–276. http://dx.doi.org/10.1016/j.molcel.2012.08.007.
42. **Nishimoto M, Fukushima A, Okuda A, Muramatsu M.** 1999. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol Cell Biol **19:**5453–5465.
43. **Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y.** 2011. Dual functions of Tet1 in transcriptional regulation in

mouse embryonic stem cells. Nature **473:**389–393. http://dx.doi.org/10.1038/nature09934.

44. **Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, et al.** 2012. The accessible chromatin landscape of the human genome. Nature **489:**75–82. http://dx.doi.org/10.1038/nature11232.
45. **Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D.** 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature **480:**490–495. http://dx.doi.org/10.1038/nature10716.
46. **Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A.** 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One **5:**e8888. http://dx.doi.org/10.1371/journal.pone.0008888.
47. **Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C.** 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell **149:**1368–1380. http://dx.doi.org/10.1016/j.cell.2012.04.027.
48. **Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES.** 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature **454:**766–770. http://dx.doi.org/10.1038/nature07107.
49. **Tsagaratou A, Rao A.** 2013. TET proteins and 5-methylcytosine oxidation in the immune system. Cold Spring Harb Symp Quant Biol **78:**1–10. http://dx.doi.org/10.1101/sqb.2013.78.020248.
50. **Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RDG.** 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature **448:**196–199. http://dx.doi.org/10.1038/nature05972.
51. **Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W, Reik W, Bertone P, Smith A.** 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell **158:**1254–1269. http://dx.doi.org/10.1016/j.cell.2014.08.029.
52. **Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, Lao K, Surani MA.** 2010. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell **6:**468–478. http://dx.doi.org/10.1016/j.stem.2010.03.015.
53. **Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A.** 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature **484:**339–344. http://dx.doi.org/10.1038/nature10960.
54. **Jin C, Lu Y, Jelinek J, Liang S, Estecio MRH, Barton MC, Issa J-PJ.** 2014. TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. Nucleic Acids Res **42:**6956–6971. http://dx.doi.org/10.1093/nar/gku372.
55. **Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PAC, Rappsilber J, Helin K.** 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature **473:**343–348. http://dx.doi.org/10.1038/nature10066.
56. **Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH.** 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell **133:**1106–1117. http://dx.doi.org/10.1016/j.cell.2008.04.043.
57. **Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hubner K, Scholer HR.** 1996. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development **122:**881–894.
58. **Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R.** 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature **494:**497–501. http://dx.doi.org/10.1038/nature11884.
59. **Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W.** 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature **473:**398–402. http://dx.doi.org/10.1038/nature10008.
60. **Foley JW, Sidow A.** 2013. Transcription-factor occupancy at HOT regions quantitatively predicts RNA polymerase recruitment in five human cell lines. BMC Genomics **14:**720. http://dx.doi.org/10.1186/1471-2164-14-720.
61. **Kvon EZ, Stampfel G, Yanez-Cuna JO, Dickson BJ, Stark A.** 2012. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. Genes Dev **26:**908–913. http://dx.doi.org/10.1101/gad.188052.112.
62. **Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, Le T, Faull KF, Jaenisch R, Tsai L-H.** 2013. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron **79:**1109–1122. http://dx.doi.org/10.1016/j.neuron.2013.08.003.
63. **Zhang RR, Cui QY, Murai K, Lim YC, Smith ZD, Jin S, Ye P, Rosa L, Lee YK, Wu HP, Liu W, Xu ZM, Yang L, Ding YQ, Tang F, Meissner A, Ding C, Shi Y, Xu GL.** 2013. Tet1 regulates adult hippocampal neurogenesis and cognition. Cell Stem Cell **13:**237–245. http://dx.doi.org/10.1016/j.stem.2013.05.006.
64. **Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, Kou X, Zhang Y, Huang H, Jiang Y, Yao C, Liu X, Lu Z, Xu Z, Kang L, Chen J, Wang H, Cai T, Gao S.** 2013. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell **12:**453–469. http://dx.doi.org/10.1016/j.stem.2013.02.005.
65. **Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA, Bruneau BG.** 2012. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell **151:**206–220. http://dx.doi.org/10.1016/j.cell.2012.07.035.
66. **Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS.** 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol **9:**R137. http://dx.doi.org/10.1186/gb-2008-9-9-r137.
67. **Kumaki Y, Masaaki Oda M, Okano M.** 2008. QUMA: quantification tool for methylation analysis. Nucleic Acids Res **36:**W170–W175. http://dx.doi.org/10.1093/nar/gkn294.